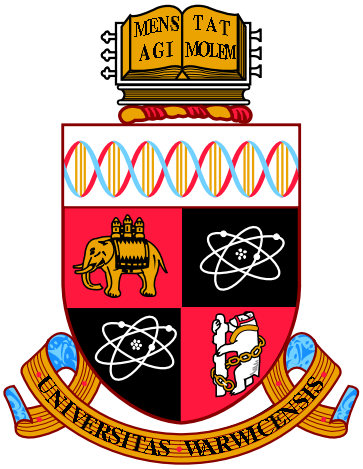

Interacting Particle Approximations of Feynman-Kac Measures for Continuous-Time Jump Processes

by

Letizia Angeli

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Mathematics Institute

September 2020

Contents

Acknowledgments

First, I would like to express my deepest gratitude to my supervisors Stefan Grosskinsky and Adam Johansen, for their patience and continuous support during my PhD. I am grateful for the time they invested in me and for guiding me throughout this whole journey.

I would also like to thank all my colleagues from Masdoc and from the enrichment program at the Alan Turing Institute for the very friendly and inspirational working environment.

Last but not least, I thank my family, my closest friends and my partner for their love, support and encouragement.

Declarations

This work has been composed by myself and has not been submitted for any other degree or professional qualification. Except where it is stated otherwise, all the results of this thesis are, to the best of my knowledge, original and my own work.

Chapter 2 presents the basic theory of Feller processes and Feynman-Kac models. For this introductory chapter, I do not claim any originality in the material itself, but only in the presentation, and in the choice of contents.

The work in Sections 3.1-3.2, Sections 4.1-4.2, and Sections 5.1-5.4 has been submitted for publication (Angeli et al., 2020), in collaboration with Stefan Grosskinsky and Adam Johansen. I stated and proved all the results presented here and I wrote the first version of the paper, which was subsequently much improved with Stefan's and Adam's input.

The work in Section 5.5 has been published in (Angeli et al., 2019), in collaboration with Stefan Grosskinsky, Adam Johansen and Andrea Pizzoferrato. My contribution to this paper was in proving the theoretical results, whereas I did not contribute to the numerical experiments.

Finally, the results presented in Section 3.3 and in Section 4.3 are original and they have not been submitted for publication.

Abstract

The cloning algorithm has been introduced in the theoretical physics literature as a numerical procedure to study the large deviations of additive path functionals of stochastic processes. Due to its broad applicability, the convergence properties of the cloning algorithm have recently become a subject of research interest with only partial progress in continuous time. In this work, we derive rigorous convergence results by providing a novel interpretation of the cloning algorithm based on Feynman-Kac models and their particle approximations. We then adapt already established convergence results for mean field particle approximations to a broader class of interacting particle approximations, which includes cloning algorithms. This way, we obtain bias and L^p error bounds, with order of convergence given respectively by $1/N$ and $1/\sqrt{N}$, and a Central Limit Theorem. Finally, we show how to apply these results to the study of large deviations of additive path functionals for Markov processes and, in particular, how to construct efficient interacting particle approximations for estimating the scaled cumulant generating function. Our results apply to a vast class of jump processes on locally compact state spaces, and do not involve any time discretization in contrast to previous approaches. This also provides a rigorous framework that can be used to explore the various degrees of freedom in the design of interacting particle approximations and to improve the efficiency of the algorithms.

Chapter 1

Introduction

The study of large deviations for additive path functions of stochastic processes arises in many physical applications such as molecular dynamics, chemical reactions and current fluctuations of non-equilibrium lattice gas models. Unfortunately, when a system is complex enough, it ceases to be feasible to simulate repeatedly the true dynamics to observe a rare event. One solution is to generate the atypical trajectories in a controlled way.

A widely-used class of numerical procedures for generating rare events efficiently are importance sampling methods based on *cloning algorithms* (Giardinà et al., 2006; Lecomte and Tailleur, 2007), which are used to evaluate numerically the *scaled cumulant generating function* (SCGF) of additive observables in Markov processes, a quantity which plays an essential role in the study of large deviations of non-equilibrium systems. Cloning algorithms combine importance sampling with a stochastic selection mechanism, based on classical ideas of evolutionary algorithms (Anderson, 1975; Grassberger, 2002): a fixed size population of copies of the original system evolves in parallel, subject to cloning or killing in such a way as to favour the realisation of atypical trajectories. Various variants of this approach are now applied to different physical systems of interest, including current fluctuations of non-equilibrium lattice gas models (Hurtado and Garrido, 2009; Hurtado et al., 2014; Nemoto et al., 2019; Chleboun et al., 2018), dynamical phase transitions (Pérez-Espigares and Hurtado, 2019), turbulent flows (Lestang et al., 2020), glassy dynamics (Pitard et al., 2011), heat waves in climate models (Ragone et al., 2018), and pressure of the edge-triangle model (Giardinà et al., 2020). Due to its widespread applications, the mathematical justification and convergence properties of the cloning algorithm have recently become a subject of research interest with only partial progress. Formal approaches so far are based on a branching process

interpretation of the algorithm in discrete time (Nemoto et al., 2017), with limited and mostly numerical results in continuous time (Hidalgo et al., 2017; Tchernookov and Dinner, 2010; Nemoto et al., 2016; Brewer et al., 2018).

In this thesis, we provide a novel interpretation of the cloning algorithm through Feynman-Kac models and their particle approximations (Del Moral, 2004; Del Moral and Miclo, 2000), which is itself an established approach to understand sequential Monte Carlo (SMC) methods and particle filtering. Feynman-Kac models were originally introduced in the 1940s (Kac, 1949) to express the semigroup of a quantum particle evolving in a potential in terms of a functional-path integral formula. In turn, Kac was inspired by Feynman (1949), who provided a heuristic connection between the Schrödinger equation and path integration theory. The key idea behind Feynman-Kac models is to enter the effects of a potential into the distribution of the paths of a stochastic process. One of the main advantages of this interpretation is that it is possible to construct explicitly N-particle systems which converge to the associated Feynman-Kac model as the size of the system tends to infinity. In particular, the Feynman-Kac representation provides natural semigroup and martingale techniques to analyze the asymptotic behaviour of the associated interacting particle approximation. These considerations have inspired us to find the Feynman-Kac interpretation of the SCGF and to understand the cloning algorithm as a particular interacting particle approximation of Feynman-Kac formulae.

More formally, we consider a pure jump Markov process X_t on a locally compact Polish state space E, with initial distribution μ_0 and with associated expectation denoted by $\mathbb{E}_{\mu_0}$. This setting covers in particular any finite-state Markov chain or stochastic interacting system on a bounded domain (e.g. finite lattice) with bounded total mass. We are interested in studying the large deviations associated with a time-additive observable A_T taken to be a real measurable function of the paths of X_t over the time interval $[0,\ T]$, in particular our aim is to provide an estimate of the scaled cumulant generating function (SCGF),

$$\Lambda(k) := \lim_{T\to\infty} \frac{1}{T} \log\ \mathbb{E}_{\mu_0}\left[e^{kTA_T}\right] , \quad k \in \mathbb{R} .$$

Under general conditions, the SCGF can be rewritten in terms of asymptotics of normalised t-marginal Feynman-Kac measures μ_t on E which solve a non-linear evolution equation in the form

$$\frac{d}{dt}\,\mu_t(f) \ = \ \mu_t\Big(\widehat{\mathcal{L}}_k f + \mathcal{V}_k f - \mu_t(\mathcal{V}_k) f\Big) ,$$

for any bounded continuous real-valued function $f \in \mathcal{C}_b(E)$, where $\widehat{\mathcal{L}}_k$ is a probability-preserving modified version of the infinitesimal generator of the original jump process and $\mathcal{V}_k \in \mathcal{C}(E)$ is a potential function related to the observable A_T. A key property of Feynman-Kac models is that μ_t can be interpreted as the law of a non-linear Markov process associated to a nonlinear probability generator in the form $\widehat{\mathcal{L}}_k + \widetilde{\mathcal{L}}_{\mu_t}$, known as *McKean interpretation* (Del Moral, 2004). The non-linear part of the generator, $\widetilde{\mathcal{L}}_{\mu_t}$, is not unique and the accuracy of the corresponding particle approximation also depends on the choice of the McKean interpretation. The rates associated to $\widetilde{\mathcal{L}}_{\mu_t}$ depend on the distribution of μ_t itself, which in general is not known a priori. A standard approach is to sample such processes through particle approximations, which consist in running in parallel N copies (or 'particles') $\underline{X}_t^N := (X_t^1, \ldots, X_t^N)$ of the process, where the interactions between particles are introduced by 'selection' events consisting in replacing one or more particles with copies of another particle appropriately chosen from the population. Then, μ_t is approximated by the empirical distribution of the realisations

$$m(\underline{X}_t^N) := \frac{1}{N} \sum_{i=1}^{N} \delta_{X_t^i} \ .$$

The most basic particle approximation is given by *mean field particle systems* (Del Moral, 2013, Section 5.4), which consist in simply running the McKean dynamic in parallel on each of the particles. The study of the asymptotic behaviour of these models is based on semigroup and martingale techniques (Del Moral and Miclo, 2000, 2007; Rousset, 2006; Cérou et al., 2016). This literature generally considers diffusive dynamics and relies upon approximate time-discretisations.

One of the two main contributions in this thesis is the adaptation of well-established convergence results for mean field particle approximations, in the context of jump processes, to a broader class of interacting particle approximations, which includes cloning algorithms (Angeli et al., 2020). In particular, we provide general assumptions on the infinitesimal generator and carré du champ of an interacting particle approximation $\underline{X}_t^N$ which guarantee that the empirical distribution $m(\underline{X}_t^N)$ converges uniformly in time to μ_t in L^p and almost surely in the weak topology, as the size N of the population goes to infinity. Moreover, we show that the order of convergence of the bias and L^p error is given respectively by $1/N$ and $1/\sqrt{N}$, as for mean field approximations. In a similar fashion, we also provide a central limit theorem and explicit formulas for the asymptotic variance of the algorithms. The framework we develop here can be used to explore several degrees of freedom

in the design of interacting particle approximations that can be used to improve performance.

The second major contribution is about the study of convergence of the cloning algorithm (Angeli et al., 2019). We provide an infinitesimal description of the cloning algorithm through Markov generators and the associated carré du champ to show that the cloning algorithm is part of the above class of interacting particle approximations and, hence, the aforementioned limit theorems apply to cloning too. In contrast to previous work in the context of cloning algorithms (Nemoto et al., 2017; Hidalgo et al., 2017), our mathematical approach does not require a time discretization and works in the very general setting of a pure jump Markov process on a locally compact state space. This covers in particular any finite-state Markov chain or stochastic particle systems on finite lattices with bounded total mass.

The thesis is structured as follows. In Chapter 2, we introduce various mathematical concepts used throughout this work, such as Markov semigroups, infinitesimal generators and the carré du champ operator. In the second part of the Chapter, we introduce general Feynman-Kac models associated to pure jump Markov processes and show that they can be interpreted as the law of a non-linear Markov process, known as a McKean interpretation (Del Moral, 2004). In Chapter 3, we describe a general class of interacting particle approximations for Feynman-Kac models - which includes cloning and mean field particle methods - and we adapt the convergence results for mean field particle approximations (Del Moral and Miclo, 2000; Rousset, 2006) to this more general class of interacting particle approximations. In particular, we establish general conditions for rigorously studying the L^p error and asymptotic variance of estimators associated with a given interacting particle approximation. In Chapter 4, we introduce mean field particle approximations and cloning algorithms, and we provide a discussion on their convergence properties. We conclude the chapter by presenting a third interacting particle approximation, the *resampling algorithm*, which does not satisfy the assumptions for studying the convergence of the algorithm, and we show in particular that the convergence results in Chapter 3 do not hold. In Chapter 5, we introduce large deviations and scaled cumulant generating functions (SCGF) of additive observables for pure jump Markov processes and discuss how the results presented in Chapter 3 can be applied to estimate the SCGF. We conclude the chapter by illustrating how to apply interacting particle approximations to current fluctuations of lattice gases using the inclusion process as an example. We conclude with a short discussion in Chapter 6.

Chapter 2

Preliminaries

In this chapter, we introduce the setting of Markov pure jump processes and present the general theory for continuous-time Feynman-Kac models. The definitions and results presented in Section 2.1 are based on Liggett (2010); Kutner and Masoliver (2017), whereas the content presented in Section 2.2 is based on Del Moral (2004); Del Moral and Miclo (2000) and references therein.

2.1 The setting

2.1.1 Feller processes

Throughout this dissertation, E is a locally compact Polish state space. The measurable structure on E is given by the Borel σ-algebra $\mathcal{B}(E)$. We denote by $\mathcal{P}(E)$ the set of probability measures on E and by $\mathcal{C}_b(E)$ the space of bounded continuous real-valued functions on E, with the uniform norm

$$||f|| := \sup_{x \in E} |f(x)|,$$

which makes $\mathcal{C}_b(E)$ a Banach space.

The *path space* of the processes considered is

$$\Omega = D([0,\infty),\, E) := \{\omega : [0,\infty) \to E \text{ càdlàg}\},$$

where *càdlàg* stands for right-continuous functions with left limits. The measurable structure on Ω is given by the Borel σ-algebra $\mathcal{F}$ induced by the Skorohod topology. The right-continuous filtration $(\mathcal{F}_t)_{t\geq 0}$ on Ω is such that the σ-algebra $\mathcal{F}_t$ is the smallest such that the mapping $\omega \mapsto \omega(t)$, with $\omega \in \Omega$, is $\mathcal{F}_t$-measurable for each $t \geq 0$.

A continuous-time stochastic process on E is denoted by $(X_t)_{t\geq 0}$, i.e. a family of random variables X_t. The canonical construction of the random variables is $X_t(\omega) = \omega(t)$. In general, the process $(X_t)_{t\geq 0}$ is characterised by a probability measure $\mathbb{P}$ on the filtered measure space $(\Omega, \mathcal{F}, \mathcal{F}_t)$ with associated expectation denoted by $\mathbb{E}$. If we want to stress a particular initial condition $x \in E$ of the process we write $\mathbb{P}_x$ and $\mathbb{E}_x$. With ω drawn from this measure, the function $t \mapsto X_t(\omega)$ is called sample path. In the following we restrict ourselves to Feller processes, which are defined as follows.

Definition 2.1.1 (Feller Process)**.** A *Feller process* on E consists of a collection of probability measures $(\mathbb{P}_x)_{x\in E}$ on $(\Omega, \mathcal{F})$ and a right-continuous filtration $(\mathcal{F}_t)_{t\geq 0}$ on Ω with respect to which the random variables X_t are adapted, satisfying

- $\mathbb{P}_x(X_0 = x) = 1$,
- (Feller property) the mapping $x \mapsto \mathbb{E}_x[f(X_t)]$ is in $\mathcal{C}_b(E)$ for all $f \in \mathcal{C}_b(E)$ and $t \geq 0$,
- (Markov property) $\mathbb{E}_x[Y \circ \theta_s \,|\, \mathcal{F}_s] = \mathbb{E}_{X_s}[Y]$, $\mathbb{P}_x$-almost surely, for all $x \in E$ and all bounded measurable Y on Ω. Here $\theta_s\, \omega_t := \omega_{t+s}$ denotes a time shift.

The Feller property allows to consider processes with general initial distribution $\mu_0 \in \mathcal{P}(E)$ via $\mathbb{P}_{\mu_0} := \int_E \mathbb{P}_x\, \mu_0(dx)$. One advantage of Feller processes is that they can be described via semigroups, infinitesimal generators or through martingales.

Definition 2.1.2 (Semigroup)**.** A family of continuous linear operators $\big(P(t)\big)_{t\geq 0}$ on $\mathcal{C}_b(E)$ is called *semigroup* if it satisfies the following properties:

- $P(0)f = f$ for all $f \in \mathcal{C}_b(E)$,
- $\lim_{t\to 0_+} P(t)f = f$, for all $f \in \mathcal{C}_b(E)$,
- $P(s+t)f = P(s)P(t)f$, for all $f \in \mathcal{C}_b(E)$ and $s, t \geq 0$,
- $P(t)f \geq 0$, for all non-negative $f \in \mathcal{C}_b(E)$ and $t \geq 0$.

If, in addition, $P(t)1 = 1$ for each $t \geq 0$, then $\big(P(t)\big)_{t\geq 0}$ is called *probability semigroup*.

Definition 2.1.3 (Generator)**.** An *infinitesimal generator* is a linear operator $\mathcal{L}$ on $\mathcal{C}_b(E)$ (possibly unbounded), with domain $\mathcal{D}_\mathcal{L}$ and range $\mathcal{R}(\mathcal{L})$, satisfying the following properties:

- $\mathcal{D}_{\mathcal{L}}$ is dense in $\mathcal{C}_b(E)$,
- if $f \in \mathcal{D}_{\mathcal{L}}$, $\lambda \geq 0$ and $g := f - \lambda\mathcal{L}(f)$, then

$$\inf_{x\in E} f(x) \geq \inf_{x\in E} g(x),$$

- $\mathcal{R}(\mathbb{1} - \lambda\mathcal{L}) = \mathcal{C}_b(E)$, for sufficiently small $\lambda > 0$, where $\mathbb{1}$ denotes the identity function.

If in addition $\mathcal{L}(1) = 0$, then $\mathcal{L}$ is called *probability generator*.

The following two results underline the connection of a Feller process with its infinitesimal generator, through its probability semigroup.

Proposition 2.1.4. *Given a Feller process X_t, define*

$$P(t)f(x) \;=\; \mathbb{E}_x f(X_t),$$

for $f \in \mathcal{C}_b(E)$. Then $P(t)$ is a probability semigroup on $\mathcal{C}_b(E)$.

Proof. See Liggett (2010), Theorem 3.15. □

Remark. We stress that, if E is not compact and thus $\mathcal{C}_b(E)$ is not separable, the converse of Proposition 2.1.4 doesn't hold in general, i.e. given a probability semigroup $P(t)$ on $\mathcal{C}_b(E)$, the existence of a Feller process X_t satisfying $\mathbb{E}_x[f(X_t)] = P(t)f(x)$ is not ensured (see Liggett, 2010, Theorem 3.26).

Theorem 2.1.5 (Hille-Yosida)**.** *Suppose that $P(t)$ is a probability semigroup and define $\mathcal{L}$ by*

$$\mathcal{L}(f) := \lim_{t\to 0} \frac{P(t)f - f}{t} \tag{2.1}$$

on

$$\mathcal{D}_{\mathcal{L}} := \{f \in \mathcal{C}_b(E) \,|\, \text{the limit (2.1) exists}\}\ .$$

Then $\mathcal{L}$ is a probability generator. Furthermore, the following statements hold:

- *if $f \in \mathcal{D}_{\mathcal{L}}$, then $P(t)f \in \mathcal{D}_{\mathcal{L}}$ for all $t \geq 0$, and it is a continuously differentiable function on t and satisfies*

$$\frac{d}{dt}P(t)f \;=\; P(t)\mathcal{L}(f) \;=\; \mathcal{L}(P(t)f);$$

- *for $f \in \mathcal{C}_b(E)$ and $t > 0$,*

$$\lim_{n\to\infty} \left(\mathbb{1} - \frac{t}{n}\mathcal{L} \right)^{-n} f \;=\; P(t)f.$$

Proof. See Liggett (2010), Theorem 3.16. □

To study the fluctuation of a Feller process, it is useful to introduce the carré du champ associated to a generator.

Definition 2.1.6 (carré du champ)**.** Given a probability generator $\mathcal{L}$ on $\mathcal{D}_\mathcal{L}$, the *carré du champ* associated to $\mathcal{L}$ is the bilinear operator on $\mathcal{D}_\mathcal{L} \times \mathcal{D}_\mathcal{L}$ given by

$$\Gamma_\mathcal{L}(\gamma, \varphi) := \mathcal{L}(\gamma \cdot \varphi) - \gamma \cdot \mathcal{L}(\varphi) - \varphi \cdot \mathcal{L}(\gamma) \;,$$

for any $\gamma, \varphi \in \mathcal{D}_\mathcal{L}$.

Lemma 2.1.7. *Let X_t be a Feller process adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$, with semigroup $P(t)$ and generator $\mathcal{L}$ with domain $\mathcal{D}_\mathcal{L} \subseteq \mathcal{C}_b(E)$ and let $\mathcal{D}_\mathcal{L} \times \mathcal{C}_b^1(\mathbb{R}^+)$ denote the space of continuous real-valued functions $\varphi_\cdot$ on $E \times \mathbb{R}^+$, such that $\varphi_t \in \mathcal{D}_\mathcal{L}$, for any $t \geq 0$, and $t \mapsto \varphi_t(x)$ is a bounded continuous function with bounded continuous first derivative, for any $x \in E$.*

For any $\varphi_\cdot \in \mathcal{D}_\mathcal{L} \times \mathcal{C}_b^1(\mathbb{R}^+)$, the process

$$\mathcal{M}_t(\varphi_\cdot) \;:=\; \varphi_t\big(X_t\big) \;-\; \varphi_0\big(X_0\big) \;-\; \int_0^t \big(\partial_s\varphi_s \;+\; \mathcal{L}(\varphi_s)\big)\big(X_s\big)\, ds \tag{2.2}$$

is a martingale relative to $\mathbb{P}_x$, for every $x \in E$, with predictable quadratic variation given by

$$\big\langle \mathcal{M}_\cdot(\varphi_\cdot) \big\rangle_0^t \;=\; \int_0^t \Gamma_\mathcal{L}(\varphi_s,\, \varphi_s)\big(X_s\big)\; ds \;. \tag{2.3}$$

Proof. The proof that $\mathcal{M}_t(\varphi_\cdot)$ is a martingale is a simple generalisation of Liggett (2010, Theorem 3.32) to time-dependent functions. Indeed, for any $\gamma_\cdot \in \mathcal{D}_\mathcal{L} \times \mathcal{C}_b^1(\mathbb{R}^+)$, the expectation of the process (2.2) can be written as

$$\begin{aligned} \mathbb{E}_x\big[\mathcal{M}_t(\gamma_\cdot)\big] \;&=\; P(t)\gamma_t(x) - \gamma_0(x) \;-\; \int_0^t P(s)\big(\partial_s\gamma_s + \mathcal{L}(\gamma_s)\big)(x)\, ds \\ &=\; 0 \;, \end{aligned}$$

since $P(s)\big(\partial_s\gamma_s + \mathcal{L}(\gamma_s)\big) \;=\; \frac{d}{ds}P(s)(\gamma_s)$, by Hille-Yosida (Theorem 2.1.5). Moreover,

$\mathbb{E}_x\big[\big|\mathcal{M}_t(\varphi_\cdot)\big|\big] < \infty$ for any $t \in \mathbb{R}^+$, and

$$\begin{aligned}
&\mathbb{E}_x\Big[\mathcal{M}_t(\varphi_\cdot)\,\Big|\,\mathcal{F}_u\Big] \\
&\quad = \mathcal{M}_u(\varphi_\cdot) + \mathbb{E}\Big[\varphi_t(X_t) - \varphi_u(X_u) - \int_u^t \big(\partial_s\varphi_s + \mathcal{L}(\varphi_s)\big)\big(X_s\big)\,ds\,\Big|\,\mathcal{F}_u\Big]\,,
\end{aligned}$$

for any $0 \le u \le t$. Substituting $\gamma_s = \varphi_{s+u}$, we obtain

$$\begin{aligned}
&\mathbb{E}_x\Big[\varphi_t(X_t) - \varphi_u(X_u) - \int_u^t \big(\partial_s\varphi_s + \mathcal{L}(\varphi_s)\big)\big(X_s\big)\,ds\,\Big|\,\mathcal{F}_u\Big] \\
&\quad = \mathbb{E}_{X_u}\Big[\gamma_{t-u}(X_{t-u}) - \gamma_0(X_0) - \int_0^{t-u} \big(\partial_s\gamma_s + \mathcal{L}(\gamma_s)\big)\big(X_s\big)\,ds\Big] \\
&\quad = \mathbb{E}_{X_u}\Big[\mathcal{M}_{t-u}(\gamma_\cdot)\Big] = 0.
\end{aligned}$$

For proving the second part of the statement, we have to show that

$$\mathcal{M}_t(\varphi_\cdot)^2 - \int_0^t \Gamma_{\mathcal{L}}(\varphi_s,\,\varphi_s)\big(X_s\big)\,ds$$

is a martingale. Note that

$$\begin{aligned}
\mathcal{M}_t(\varphi_\cdot)^2 &= \int_0^t 2\mathcal{M}_s(\varphi_\cdot)d\mathcal{M}_s(\varphi_\cdot) \\
&= \int_0^t 2\,\big(\varphi_s(X_s) - \varphi_0(X_0)\big)\,d\mathcal{M}_s - 2\int_0^t \Big(\int_0^s \big(\partial_u + \mathcal{L}\big)\varphi_u(X_u)\,du\Big)\,d\mathcal{M}_s\,,
\end{aligned}$$

and

$$\begin{aligned}
&\int_0^t 2\,\big(\varphi_s(X_s) - \varphi_0(X_0)\big)\,d\mathcal{M}_s \\
&\quad = \big(\varphi_t(X_t) - \varphi_0(X_0)\big)^2 - \int_0^t 2\,\big(\varphi_s(X_s) - \varphi_0(X_0)\cdot\big(\partial_s\varphi_s + \mathcal{L}(\varphi_s)\big)\,ds \\
&\quad = \int_0^t \Gamma_{\mathcal{L}_s}(\varphi_s,\,\varphi_s)(X_s)\,ds + \mathcal{M}_t\big(\varphi_\cdot^2\big) - 2\varphi_0(X_0)\cdot\mathcal{M}_t(\varphi_\cdot)\,,
\end{aligned}$$

by a simple computation. Therefore,

$$\begin{aligned}
&\mathcal{M}_t(\varphi_\cdot)^2 - \int_0^t \Gamma_{\mathcal{L}}(\varphi_s,\,\varphi_s)\big(X_s\big)\,ds \\
&\quad = \mathcal{M}_t\big(\varphi_\cdot^2\big) - 2\varphi_0(X_0)\cdot\mathcal{M}_t(\varphi_\cdot) - 2\int_0^t \Big(\int_0^s \big(\partial_u + \mathcal{L}\big)\varphi_u(X_u)\,du\Big)\,d\mathcal{M}_s\,,
\end{aligned}$$

which is a martingale, as the integral of a progressively measurable process with

respect to a martingale is itself a martingale.

□

2.1.2 Pure jump Markov processes

Under the setting presented in Section 2.1.1, we introduce the pure jump Markov processes, which will be the main object of our study.

Definition 2.1.8. A **pure jump Markov process** is a right-continuous Feller process such that there exists a sequence of strictly increasing stopping times $(T_n)_{n\geq 0}$ such that $T_0 = 0$, X_t is constant on the interval $[T_n, T_{n+1})$ and $X_{T_n^-} \neq X_{T_n}$ for every $n \geq 0$.

To describe these jumps, it is usual to introduce the *escape rate function* $\lambda(x)$, such that $\lambda(x)dt + o(dt)$ is the probability that X_t undergoes a jump during $[t, t+dt]$ starting from the state $X_t = x$. More formally, the jumping times $(T_n)_{n\geq 0}$ are given by the recursive formulae

$$T_{n+1} := \inf\left\{ t \geq T_n \,|\, \int_{T_n}^{t} \lambda(X_s)\, ds \geq e_n \right\} ,$$

where $T_0 = 0$ and $(e_n)_{n\geq 0}$ stands for a sequence of i.i.d. exponential random variables with unit parameter. Moreover, when a jump occurs, X_{t+dt} is then distributed with the *probability kernel* $p(x, dy)$. The path of a pure jump process can thus be represented by the sequence of visited states in E, together with the sequence of waiting times in those states, so that the space of paths is equivalent to $[E \times (0,\infty)]^{\mathbb{N}}$.

Assumption 2.1.9. From now on, we assume:

- $\lambda(\cdot) : E \to (0,\infty)$ to be a strictly positive, bounded and continuous function;
- $x \mapsto p(x, A)$ is a continuous function for every $A \in \mathcal{B}(E)$.

Under Assumption 2.1.9, the overall transition rate is

$$W(x, dy) := \lambda(x)\, p(x, dy) ,$$

for $(x, y) \in E^2$ and the pure jump process is a Feller process with probability generator given by

$$\mathcal{L}(f)(x) = \int_E W(x, dy)\, \big(f(y) - f(x)\big) , \quad \forall f \in \mathcal{C}_b(E),\, x \in E, \tag{2.4}$$

with domain $\mathcal{D}_{\mathcal{L}} = \mathcal{C}_b(E)$ (see Ethier and Kurtz, 2009, p. 162).

Along with jump processes on continuous compact space such as continuous-time random walks (see e.g. Kutner and Masoliver, 2017), this setting includes in particular finite-state continuous-time Markov chains. Typical examples we have in mind are given by stochastic particle systems on $E = S^{\Lambda}$; in this context, if S is finite then E is compact even if Λ is countably infinite, whereas if S is countably infinite and Λ is finite then E is locally compact. Classical examples are spin systems with $S = \{-1, 1\}$ or exclusion processes with $S = \{0, 1\}$, in which particles can jump only onto empty sites (see e.g. Liggett, 2012).

The next result extends the domain of the infinitesimal generator of a pure jump process to a broader class of (possibly unbounded) continuous functions. This justifies the discussion in Section 5.4, where we make use of unbounded continuous functions to construct an unbiased estimator for unnormalized Feynman-Kac measures based on the cloning factor.

Lemma 2.1.10. *Let X_t be a pure jump process satisfying Assumption 2.1.9. Let $f \in \mathcal{C}(E)$ be a (possibly unbounded) continuous function on E such that*

$$\int_E \big|f(y) - f(x)\big|\, p(x,\, dy) \;\leq\; c \cdot \max\{1,\, f(x)\}\ ,$$

for all $x \in E$. If $f(X_0)$ is integrable, then also $f(X_t)$ is integrable, moreover the process

$$\mathcal{M}_t(f) \;:=\; f(X_t) \;-\; f(X_0) \;-\; \int_0^t \mathcal{L}(f)(X_s)\, ds \tag{2.5}$$

is a martingale and Dynkin's formula

$$\mathbb{E}\big[f(X_t)\big] \;=\; \mathbb{E}\big[f(X_0)\big] \;+\; \int_0^t \mathbb{E}\big[\mathcal{L}(f)(X_s)\big]\, ds$$

holds.

Proof. See (Hamza and Klebaner, 1995, Theorem 2). □

For example, applying Lemma 2.1.10 to pure jump processes on a metric space (E, d) with uniformly bounded jumps (i.e. there exists $k > 0$ such that $p(x,\, dy) = 0$ for any $x, y \in E$ with $d(x, y) > k$) and considering the function $f(x) = x$ (or similarly $f(x) = \log x$), we obtain that the process $\mathcal{M}_t(f)$ (2.5) is a martingale.

2.2 Feynman-Kac models

Feynman-Kac models arise in a wide variety of contexts, such as particle physics, biology, nonlinear filtering or financial mathematics. The basic idea is to enter the effects of a potential into the evolution of a Markov process. More precisely, Feynman-Kac measures represent the semigroup of a given Feller process X_t on E weighted by a potential function $\mathcal{V} \in \mathcal{C}_b(E)$. In Chapter 5 we will show, as an application, how to interpret large deviations for additive path functionals of stochastic processes via Feynman-Kac models. One main advantage of Feynman-Kac models is that they can be interpreted as genetic type particle models, with mutation-selection transitions, as we illustrate in the following chapters.

Lemma 2.2.1. *Consider a potential function $\mathcal{V} \in C_b(E)$ and the tilted generator*

$$\mathcal{L}^{\mathcal{V}}(f)(x) := \mathcal{L}(f)(x) + \mathcal{V}(x)f(x) \quad \text{defined for all } f \in C_b(E) \ . \tag{2.6}$$

Then the family of operators $\big(P^{\mathcal{V}}(t) : t \geq 0\big)$ with $P^{\mathcal{V}} : C_b(E) \to C_b(E)$, defined as the solution to the backward equation

$$\frac{d}{dt}P^{\mathcal{V}}(t)f = \mathcal{L}^{\mathcal{V}}\big(P^{\mathcal{V}}(t)f\big) \quad \text{with} \quad P^{\mathcal{V}}(0)f = f \tag{2.7}$$

for all $f \in C_b(E)$, forms a non-conservative semigroup, the so-called Feynman-Kac semigroup, and $\mathcal{L}^{\mathcal{V}}$ is its infinitesimal generator in the sense of the Hille-Yosida Theorem.

Proof. See (Liggett, 2010, Theorem 3.47). □

The semigroup $P^{\mathcal{V}}(t)$ can be also expressed in terms of the original process X_t and potential $\mathcal{V}$ via the usual formula

$$P^{\mathcal{V}}(t)f(x) \ = \ \mathbb{E}_x\Big[f(X_t) \cdot \exp\Big(\int_0^t \mathcal{V}(X_s)\, ds\Big)\Big] \ .$$

In order to control the asymptotic behaviour of $P^{\mathcal{V}}(t)$, we make the following assumption, which closely resembles (Rousset, 2006, Assumption 1) on asymptotic stability.

Assumption 2.2.2 (Asymptotic Stability)**.** The spectrum of $\mathcal{L}^{\mathcal{V}} = \mathcal{L} + \mathcal{V}$ (2.6) is bounded by a principal eigenvalue λ_0. Moreover, λ_0 is associated to a positive eigenfunction $r \in \mathcal{C}_b(E)$ and an eigenmeasure $\mu_\infty \in \mathcal{P}(E)$. Finally, there exist

constants $\alpha > 0$ and $\rho \in (0,1)$ such that

$$\left\| e^{-t\lambda_0} P^{\mathcal{V}}(t) f(\cdot) - \mu_\infty(f) \right\| \leq \|f\| \cdot \alpha \rho^t \ , \tag{2.8}$$

for every $t \geq 0$ and $f \in \mathcal{C}_b(E)$.

Asymptotic stability is for example guaranteed for all irreducible, finite-state continuous-time Markov chains which necessarily have a spectral gap. For alternative sufficient conditions implying asymptotic stability in a more general context including continuous state spaces, see Section 2.4.

We introduce the measures ν_{t,μ_0} for any general initial distribution $\mu_0 \in \mathcal{P}(E)$ and $t \geq 0$, defined by

$$\nu_{t,\mu_0}(f) := \mu_0\big(P^{\mathcal{V}}(t) f\big), \tag{2.9}$$

for any $f \in \mathcal{C}_b(E)$. In the literature (Del Moral, 2004), ν_t is known as the *unnormalised t-marginal Feynman-Kac measure*. Applying Lemma 2.2.1, we can see that ν_t solves the evolution equation

$$\frac{d}{dt}\nu_{t,\mu_0}(f) = \nu_{t,\mu_0}\big(\mathcal{L}^{\mathcal{V}}(f)\big) = \nu_{t,\mu_0}\big(\mathcal{L}(f) + \mathcal{V} \cdot f\big), \tag{2.10}$$

for any $f \in \mathcal{C}_b(E)$, $t \geq 0$ and $\mu_0 \in \mathcal{P}(E)$. The measures with which one can most naturally associate a process are the corresponding *normalised t-marginal Feynman-Kac measures* in $\mathcal{P}(E)$,

$$\mu_{t,\mu_0}(f) := \frac{\nu_{t,\mu_0}(f)}{\nu_{t,\mu_0}(1)}, \tag{2.11}$$

defined for any $t \geq 0$ and $f \in \mathcal{C}_b(E)$.

Lemma 2.2.3. *Under Assumption 2.2.2 on asymptotic stability, there exist constants $\tilde{\alpha} \geq 0$ and $0 < \rho < 1$ such that for any $f \in \mathcal{C}_b(E)$,*

$$\left| \mu_{t,\mu_0}(f) - \mu_\infty(f) \right| \leq \|f\| \cdot \tilde{\alpha} \rho^t \ , \tag{2.12}$$

for any $t \geq 0$ and independently of the initial distribution $\mu_0 \in \mathcal{P}(E)$, where $\mu_\infty \in \mathcal{P}(E)$ is the eigenmeasure associated to the principal eigenvalue λ_0 w.r.t. the generator $\mathcal{L}^{\mathcal{V}}$. In particular μ_{t,μ_0} converges weakly to μ_∞, as $t \to \infty$.

Proof. By definition of μ_{t,μ_0} (2.11) and then by asymptotic stability (Assumption 2.2.2),

$$\frac{\mu_\infty(f) - \|f\|\,\alpha \cdot \rho^t}{1 + \alpha \cdot \rho^t} \leq \mu_{t,\mu_0}(f) = \frac{\mu_0\big(e^{-t\lambda_0} P^{\mathcal{V}}(t) f\big)}{\mu_0\big(e^{-t\lambda_0} P^{\mathcal{V}}(t) 1\big)} \leq \frac{\mu_\infty(f) + \|f\|\,\alpha \cdot \rho^t}{1 - \alpha \cdot \rho^t} \ , \tag{2.13}$$

for any $t > -\log \alpha / \log \rho$ and for some constant $\alpha > 0$. This gives the bound (2.12) for any t large enough. Increasing α accordingly to ensure that the bound holds also for small t, we obtain (2.12) for any $t \geq 0$.

□

For simplicity, in the rest of this dissertation the initial distribution μ_0 is fixed and we write μ_t (resp. ν_t) instead of μ_{t,μ_0} (resp. ν_{t,μ_0}).

Now, we want to outline the evolution of the time-marginal distribution μ_t in terms of interacting jump-type infinitesimal generators. The content presented in the rest of this section is based on Del Moral (2004, 2013); Del Moral and Miclo (2000). In this established framework it is possible to define generic Markov processes with time marginals μ_t and then use Monte Carlo sampling techniques to approximate those marginals.

Lemma 2.2.4. *For every $f \in \mathcal{C}_b(E)$ and $t \geq 0$, the normalised t-marginal μ_t (2.11) solves the non-linear evolution equation*

$$\frac{d}{dt}\mu_t(f) = \mu_t\big(\mathcal{L}(f)\big) + \mu_t(\mathcal{V}f) - \mu_t(f) \cdot \mu_t(\mathcal{V}). \tag{2.14}$$

Proof. Using the evolution equation (2.10) of ν_t, we see that

$$\begin{aligned}
\frac{d}{dt}\mu_t(f) &= \frac{d}{dt}\frac{\nu_t(f)}{\nu_t(1)} \\
&= \frac{1}{\nu_t(1)} \cdot \nu_t\big(\mathcal{L}(f) + \mathcal{V} \cdot f\big) - \frac{\nu_t(f)}{\nu_t(1)^2}\,\nu_t\big(\mathcal{L}(1) + \mathcal{V}\big) \\
&= \mu_t\big(\mathcal{L}(f)\big) + \mu_t(\mathcal{V}f) - \mu_t(f) \cdot \mu_t(\mathcal{V}) \ .
\end{aligned}$$

□

The evolution equation (2.14) results from the unique decomposition of the non-conservative generator $\mathcal{L} + \mathcal{V}$ into a conservative and a diagonal part given by the potential $\mathcal{V}$. The latter, together with the normalisation of ν_t, leads to the nonlinear second part in (2.14) which we want to rewrite to be in the form of another infinitesimal generator, that we denote by $\widetilde{\mathcal{L}}_{\mu_t}$. Since (2.14) is non-linear in μ_t, this depends itself on the current distribution such that

$$\mu\big(\widetilde{\mathcal{L}}_\mu(f)\big) \ = \ \mu(\mathcal{V}f) \ - \ \mu(f) \cdot \mu(\mathcal{V}) \ , \tag{2.15}$$

for every $\mu \in \mathcal{P}(E)$ and $f \in \mathcal{C}_b(E)$.

Moreover, the evolution equation (2.14) can be interpreted as the evolution of the $\mathrm{Law}(\overline{X}_t) = \mu_t$ of a time-inhomogeneous process $\overline{X}_t$ with a probability generator

$$\overline{\mathcal{L}}_{\mu_t} := \mathcal{L} + \widetilde{\mathcal{L}}_{\mu_t} \ . \tag{2.16}$$

In this situation, it is essential to observe that $\overline{\mathcal{L}}_{\mu_t}$ depends on the current distribution μ_t of the random state $\overline{X}_t$. The stochastic model $\overline{X}_t$, or equivalently the collection of probability generators $(\overline{\mathcal{L}}_\mu)_{\mu\in\mathcal{P}(E)}$, is known as *McKean interpretation* of the non-linear evolution equation in distribution space defined in (2.14). As will become clear later, there is a natural evolutionary interpretation of the process $\overline{X}_t$ and we call the jumps given by $\mathcal{L}$ *mutation events* and the jumps given by $\widetilde{\mathcal{L}}_{\mu_t}$ *selection events*.

The evolution of the McKean model can be described using the propagator

$$\Theta_{t,T} f(x) := \frac{P^{\mathcal{V}}(T-t)f(x)}{\mu_t\big(P^{\mathcal{V}}(T-t)1\big)} \quad \text{such that} \quad \mu_T(f) = \mu_t(\Theta_{t,T} f) \ , \tag{2.17}$$

for all $0 \le t \le T$, which follows directly from the definition of μ_t (2.11) and the semigroup characterising the time evolution for ν_t (2.7). We will show in Section 3.2.1 some key properties of the propagator $\Theta_{t,T}$.

2.3 Description of some McKean models

The choice of the non-linear generator $\widetilde{\mathcal{L}}_\mu$ satisfying (2.15) is not unique, leading to various representations in the form

$$\widetilde{\mathcal{L}}_\mu(f)(x) = \int_E \widetilde{W}(x,y)\big(f(y) - f(x)\big)\mu(dy) \ , \tag{2.18}$$

where $\widetilde{W}(x,y)\mu(dy)$ is the overall transition rate of $\widetilde{\mathcal{L}}_\mu$ and depends on the current distribution μ. In order to have the generator of a well-defined Markov process, we assume $\widetilde{W}(x,y) \ge 0$, for any $x, y \in E$.

Lemma 2.3.1 (Sufficient conditions)**.** *An infinitesimal generator $\widetilde{\mathcal{L}}_\mu$ in the form (2.18) satisfies the McKean condition (2.15) if and only if*

$$\mu\big(\widetilde{W}(\cdot,x) - \widetilde{W}(x,\cdot)\big) \ = \ \mathcal{V}(x) - \mu(\mathcal{V}) \ , \tag{2.19}$$

for all $\mu \in \mathcal{P}(E)$ and $x \in E$. In particular, a sufficient condition on $\widetilde{\mathcal{L}}_\mu$ for the McKean condition (2.15) to hold is

$$\widetilde{W}(y,x) \ - \ \widetilde{W}(x,y) \ = \ \mathcal{V}(x) \ - \ \mathcal{V}(y) \ , \tag{2.20}$$

for all $x, y \in E$.

Proof. It is enough to observe that

$$\begin{aligned}\mu\big(\widetilde{\mathcal{L}}_\mu(f)\big) &= \int_{E^2} \widetilde{W}(x,y)\big(f(y)-f(x)\big)\mu(dy)\mu(dx) \\ &= \int_{E^2} \big(\widetilde{W}(y,x)-\widetilde{W}(x,y)\big)\, f(x)\mu(dy)\mu(dx)\ .\end{aligned}$$

□

Lemma 2.3.1 leads to various possible McKean representations in the form (2.16)-(2.18) (see e.g. Angeli et al., 2019; Rousset, 2006), that can be characterised by the operator $\widetilde{W}$. One common choice related to algorithms in (Giardinà et al., 2006; Lecomte and Tailleur, 2007) is

$$\widetilde{W}_c(x,y) \;=\; \big(\mathcal{V}(x)-c\big)^- \;+\; \big(\mathcal{V}(y)-c\big)^+\ , \tag{2.21a}$$

where $c \in \mathbb{R}$ is an arbitrary constant and using the standard notation $a^+ = \max\{0,a\}$ and $a^- = \max\{0,-a\}$ for positive and negative part of $a \in \mathbb{R}$. The function $\mathcal{V}(x)$ can be interpreted as a fitness potential for the process. Generic choices are:

- $c = 0$ is the default and simplest choice, but is usually not optimal as shown in Proposition 4.1.3;
- $c = \mu_t(\mathcal{V})$ corresponding to the current average potential: if the system in state x is less fit than the average it jumps with rate $\big(\mathcal{V}(x)-c\big)^-$ to a new state y chosen from the distribution $\mu_t(dy)$ (so that the overall escape rate is $\big(\mathcal{V}(x)-c\big)^-\,\mu_t(dy)$), and independently, the system jumps to states fitter than the average irrespective of its current state with overall escape rate $\big(\mathcal{V}(y)-c\big)^+\,\mu_t(dy)$;
- $c = \sup_{z\in E}\mathcal{V}(z)$ or $\inf_{z\in E}\mathcal{V}(z)$, so that the selection events are independent of the fitness potential at the target state y or at the departure state x, respectively. This simplifies the implementation.

Another common choice of $\widetilde{W}$ (2.18) satisfying (2.20) is (see e.g. Del Moral, 2013, Section 5.3.1)

$$\widetilde{W}(x,\,y) \;=\; \big(\mathcal{V}(y)-\mathcal{V}(x)\big)^+\ , \tag{2.21b}$$

which is particularly interesting for implementing efficient selection dynamics as discussed in Section 4.1.2. Here every jump from this part of the generator strictly increases the fitness of the process, which is a stronger version of the previous idea

where the process on average increased its fitness above level c. The rate depends on departure state x and target state y, which is in general computationally more expensive to implement than rates in (2.21a), but can still be feasible due to simplifications in many concrete examples as illustrated by Angeli et al. (2019). A further improvement of that idea is given by

$$\widetilde{W}(x, y) = \big(\mathcal{V}(x) - \mu(\mathcal{V})\big)^- \cdot \frac{\big(\mathcal{V}(y) - \mu(\mathcal{V})\big)^+}{\mu\big((\mathcal{V} - \mu(\mathcal{V}))^+\big)} \,, \tag{2.21c}$$

which resembles a continuous-time version of selection processes which are known under the names of stochastic remainder sampling (Baker, 1985) or residual sampling (Kong et al., 1994) in discrete time. Here selection events change the process from states x of less than average fitness $\mu(\mathcal{V})$ to states y fitter than average. However, this variant is harder to implement than (2.21b) and offers only limited extra gain on selection efficiency, as illustrated in Section 5.5 for inclusion processes (see Figure 5.2).

The process $\overline{X}_t$ can be described as follows. Between the selection events, the process $\overline{X}_t$ evolves as a copy of the reference process X_t with generator $\mathcal{L}$. The rate of the selection events is given by the function

$$\widetilde{\lambda}_{\mu_t}(x) \;=\; \mu_t\big(\widetilde{W}(x, \cdot)\big) \;.$$

Moreover, when a selection event occurs, $\overline{X}_{t+dt}$ is distributed with probability kernel

$$\widetilde{p}_t(x, dy) = \frac{\widetilde{W}(x, y)}{\mu_t\big(\widetilde{W}(x, \cdot)\big)}\, \mu_t(dy) \;.$$

Independent of the particular McKean representation, the rates of the McKean process $(\overline{X}_t : t \geq 0)$ depend on the distribution μ_t itself, which is in general not known. A standard approach is to sample such processes through particle approximations (Del Moral and Miclo, 2003), as we will see in Chapter 3.

2.4 Asymptotic stability

We present sufficient conditions for asymptotic stability as given in Assumption 2.2.2, which is essential for ensuring that the process forgets exponentially fast its initial conditions and thus for guaranteeing the stability of the associated particle approximations, as shown in Chapter 3. The discussion is based on the work of Down et al. (1995); Tweedie (1994).

Definition 2.4.1. A Feller process Y_t is said to be *ϕ-irreducible* for a non-trivial measure ϕ (i.e. $\phi(E) > 0$) on $(E, \mathcal{B}(E))$, if $\mathbb{E}_x\big[\int_0^\infty \mathbb{1}_{Y_t \in A} dt\big] > 0$ for every $x \in E$ and every set $A \in \mathcal{B}(E)$ such that $\phi(A) > 0$. We simply say that Y_t is *irreducible* if it is ϕ-irreducible for some ϕ.

Definition 2.4.2. A ϕ-irreducible Feller process Y_t is called *aperiodic* if there exist a set $C \in \mathcal{B}(E)$ with $\phi(C) > 0$, a non-trivial measure η and $t, \tau > 0$ such that

$$\mathbb{P}^x(Y_t \in B) \geq \eta(B) \quad \text{and} \quad \mathbb{P}^x(Y_s \in C) > 0 \ ,$$

for all $x \in C$, $B \in \mathcal{B}(E)$ and $s \geq \tau$.

Lemma 2.4.3. *Let Y_t be a ϕ-irreducible and aperiodic Feller process on a locally compact state space E such that* $\operatorname{supp}\phi$ *has non-empty interior. Denote by $\mathcal{L}$ and $P(t)$ the associated infinitesimal generator and the semigroup, respectively. Assume that for a given function $h \in \mathcal{C}_b(E)$ such that $h \geq 1$, there exist constants $b, c > 0$ and a compact set $S \in \mathcal{B}(E)$ such that for all $x \in E$*

$$\mathcal{L}(h)(x) \leq -c \cdot h(x) + b\mathbb{1}_S(x) \ .$$

Then there exist constants $\alpha \geq 0$ and $\rho \in (0, 1)$ such that for any test function $f \in \mathcal{C}_b(E)$ and $t \geq 0$,

$$\big|P(t)f(x) - \pi(f)\big| \leq \|f\|\, h(x) \cdot \alpha\rho^t \ ,$$

for any $x \in E$, where π is the (unique) invariant measure of Y_t.

Proof. See Down et al. (1995), Theorem 5.2(c), using the fact that if a Feller process Y_t is ϕ-irreducible and $\operatorname{supp}\phi$ has non-empty interior, then every compact set is petite (See Tweedie (1994), Theorem 7.1 and Theorem 5.1). □

In the following we discuss how the spectral properties of the (pure-jump) tilted generator $\mathcal{L}^{\mathcal{V}}$ in Assumption 2.2.2 can imply asymptotic stability in the sense of (2.8).

Assumption 2.4.4. We assume that the spectrum of $\mathcal{L}^{\mathcal{V}}$ (2.6) is bounded by a greatest eigenvalue λ_0. Moreover, there exist a positive function $r \in \mathcal{C}_b(E)$, unique up to multiplicative constants, and a probability measure $\mu_\infty \in \mathcal{P}(E)$ satisfying respectively

$$\mathcal{L}^{\mathcal{V}}(r) = \lambda_0 \cdot r \ ,$$

and

$$\mu_\infty\big(\mathcal{L}^{\mathcal{V}}(f)\big) = \lambda_0 \cdot \mu_\infty(f) \quad \text{for any } f \in \mathcal{C}(E) \ .$$

Without loss of generality, we can assume $\mu_\infty(r) = 1$.

Remark. Sufficient conditions for Assumption 2.4.4 to hold can be found, for instance, in Gong and Wu (2006); Gong et al. (2001). These are of course satisfied if the original process with generator $\mathcal{L}$ is an irreducible, finite-state Markov chain, including for example stochastic particle systems on finite lattices with a fixed number of particles.

Under Assumption 2.4.4, we define the generator

$$\mathcal{L}_r^{\mathcal{V}}(f)(x) \ = \ r^{-1}(x) \cdot \mathcal{L}^{\mathcal{V}}(r \cdot f)(x) \ - \ \lambda_0 \cdot f(x) \ , \tag{2.22}$$

which is known in the literature as *Doob's h-transform* of $\mathcal{L}^{\mathcal{V}}$ (Chetrite and Touchette, 2015) or twisted Markov kernel (Whiteley and Kantas, 2017). Observe that $\mathcal{L}_r^{\mathcal{V}}(1) = 0$, so that it is a probability generator associated to a Markov process with probability semigroup defined for any $f \in \mathcal{C}_b(E)$ by

$$P_r^{\mathcal{V}}(t)f(x) := r^{-1}(x) \cdot e^{-\lambda_0} P^{\mathcal{V}}(t)(rf)(x).$$

Proposition 2.4.5 (Asymptotic stability)**.** *Assume that there exists $\varepsilon > 0$ such that the set*

$$K_\varepsilon := \big\{x \in E \,\big|\, \mathcal{V}(x) \geq \lambda_0 - \varepsilon\big\}$$

is compact. Under Assumption 2.4.4, if the initial pure jump process $(X_t : t \geq 0)$ with generator $\mathcal{L}$ is ϕ-irreducible for some ϕ for which $\operatorname{supp}\phi$ has non-empty interior, and aperiodic as defined above then (2.8) holds, i.e. there exist $\alpha > 0$ and $\rho \in (0,1)$ such that

$$\big\|e^{-\lambda_0} P^{\mathcal{V}}(t)f - \mu_\infty(f)\big\| \leq \|f\| \cdot \alpha\rho^t$$

for every $t \geq 0$ and $f \in \mathcal{C}_b(E)$.

Proof. First, note that if the initial process X_t is irreducible and aperiodic, then also the process associated to $\mathcal{L}_r^{\mathcal{V}}$ is irreducible and aperiodic. Moreover, $\mathcal{L}_r^{\mathcal{V}}$ is bounded in K_ε and $\mathcal{L}_r^{\mathcal{V}}(r^{-1}) \leq -\varepsilon\, r^{-1}$ for every $x \notin K_\varepsilon$. Therefore, the hypotheses of Lemma 2.4.3 are satisfied for the generator $\mathcal{L}_r^{\mathcal{V}}$ acting on the function $h = r^{-1}$. Thus, applying the lemma we obtain

$$\big|P_r^{\mathcal{V}}(t)f(x) - \pi(f)\big| \leq \|f\|\, r^{-1}(x) \cdot \alpha\rho^t,$$

for any $f \in \mathcal{C}_b(E)$ and $x \in E$, where $\pi(\cdot) = \mu_\infty(r\,\cdot) \in \mathcal{P}(E)$ is the invariant measure for $\mathcal{L}_r^{\mathcal{V}}$. Dividing by $r^{-1}(x)$ and substituting f with $r^{-1}f \in \mathcal{C}_b(E)$, we obtain the statement ($\|r^{-1}\| < \infty$ and can be included in the constant α).

□

Chapter 3

Interacting Particle Approximations

In this Chapter, we present a generic description for interacting particle approximations of Feynman-Kac models, which include the classical mean field version and cloning algorithms, as illustrated in Chapter 4. In particular, we provide generalised conditions for convergence as our main result. Adapting already established convergence results for mean field approximations (Del Moral and Miclo, 2000; Rousset, 2006) to a broader class of interacting particle approximations, we provide general assumptions on the infinitesimal generator and carré du champ of the interacting particle system which guarantee that the empirical distribution converges uniformly in time to μ_t in L^p and almost surely in the weak topology. Moreover, we show that the order of convergence of the L^p and systematic errors is given respectively by $1/\sqrt{N}$ and $1/N$, as for mean field approximations. We further provide a CLT and explicit formulas for the asymptotic variance of the algorithms. These results underline the several degrees of freedom in the design of the algorithms, providing a new perspective on how to optimise the implementation of sequential Monte Carlo methodologies.

3.1 General description

This section is concerned with the general description of interacting particle approximations of Feynman-Kac t-marginal measures μ_t (2.11). These particle approximations involve running, in parallel, N copies or clones $\xi_t := (\xi_t^1, \ldots, \xi_t^N) \in E^N$ of the process (called particles), and then approximating μ_t by the empirical distri-

bution $m(\xi_t)$ of the realisations. For any $\underline{x} \in E^N$ the latter is defined as

$$m(\underline{x})(dy) := \frac{1}{N}\sum_{i=1}^{N} \delta_{x_i}(dy) \ \in \mathcal{P}(E). \tag{3.1}$$

We write $\overline{L}^N$ for the infinitesimal generator of an N-particle system ξ_t and also call this an IPS generator, and denote the associated empirical distribution as

$$\mu_t^N(\cdot) := m(\xi_t)(\cdot). \tag{3.2}$$

Recalling Definition 2.1.6, we denote by

$$\Gamma_{\overline{L}^N}(\gamma,\varphi) = \overline{L}^N(\gamma\cdot\varphi) - \gamma\cdot\overline{L}^N(\varphi) - \varphi\cdot\overline{L}^N(\gamma)\ , \qquad \gamma,\varphi\in\mathcal{C}_b(E^N)\ ,$$

the standard *carré du champ* operator associated to the generator $\overline{L}^N$.

The full dynamics can be set up in various different ways such that $\mu_t^N \to \mu_t$ converges in an appropriate sense as $N\to\infty$ for any $t\geq 0$. Theoretical convergence results can be obtained under the following assumptions, which are fulfilled by standard mean field particle approximations (as shown in Section 4.1) and cloning algorithms (Section 4.2).

Assumption 3.1.1. Given a family of McKean generators $\big(\overline{\mathcal{L}}_\mu\big)_{\mu\in\mathcal{P}(E)}$ (2.16) on $\mathcal{C}_b(E)$, we assume that the sequence of particle approximations $(\xi_t : t\geq 0)$ with generators $(\overline{L}^N)_{N\in\mathbb{N}}$ on $\mathcal{C}_b(E^N)$ satisfies

$$\overline{L}^N(F)(\underline{x}) \ = \ m(\underline{x})\big(\overline{\mathcal{L}}_{m(\cdot)}(f)\big)\ , \tag{3.3a}$$

$$\Gamma_{\overline{L}^N}(F,\,F)(\underline{x}) \ = \ \frac{1}{N}\,m(\underline{x})\big(G_{m(\cdot)}(f,f)\big) \ + \ O\bigg(\frac{1}{N^2}\bigg)\ , \tag{3.3b}$$

for mean-field observables $F\in\mathcal{C}_b(E^N)$ of the form $F(\underline{x}) = m(\underline{x})(f)$, $f\in\mathcal{C}_b(E)$. Here $\big(G_\mu\big)_{\mu\in\mathcal{P}(E)}$ is a family of bilinear operators $G_\mu : \mathcal{C}_b(E)\times\mathcal{C}_b(E)\to\mathcal{C}_b(E)$ independent of the population size N, such that

$$\sup_{\mu\in\mathcal{P}(E)}\ \sup_{\|f\|\leq 1} \|G_\mu(f,f)\| < \infty\ .$$

Furthermore, we assume there exists a constant $K<\infty$ (independent of N), such that for all $N\in\mathbb{N}$, almost surely,

$$\sup_{t\geq 0}\big|\big\{i\in 1,\ldots,N : \xi_t^i \neq \xi_{t-}^i\big\}\big| \ \leq \ K\ . \tag{3.3c}$$

For the initial condition of the particle approximation we assume that

$$\xi_0^1, \dots, \xi_0^N \quad \text{are i.i.d.r.v's with distribution } \mu_0 \ . \tag{3.3d}$$

Remark. In general, the goal is to approximate $\mu_t(f)$ for a given $f \in \mathcal{C}_b(E)$, so it is natural to set up the auxiliary particle approximation in a permutation invariant way and use test functions of the form

$$F(\underline{x}) = m(\underline{x})(f) = \frac{1}{N} \sum_{i=1}^{N} f(x_i) \ .$$

For instance, in the physics literature, these functions describe mean-field observables averaged over the particle ensemble which are generally of most interest, e.g. for the estimator (5.17) of the large deviation function it is sufficient to consider such functions, as shown in Chapter 5.

To better understand the above assumptions, recall that the carré du champ of an interacting particle system is a quadratic operator associated to the fluctuations of the process, whereas the generator determines the expected behaviour of the observables $F(\xi_t)$. Thus, Assumption 3.1.1 implies that trajectories of mean-field observables in a particle approximation coincide in expectation with average trajectories of the McKean representation they are based on (3.3a), and concentrate on their expectation with diverging N (3.3b). Condition (3.3c) assures that at any given time only a bounded number of particles can change their state, which is a mild technical assumption, necessary for applying Lemma 3.2.1 in the proof of the L^p error estimates.

Theorem 3.1.2. *Consider a sequence of particle approximations satisfying Assumption 3.1.1 with empirical distributions μ_t^N (3.2). Under Assumption 2.2.2 on asymptotic stability, for every $p \geq 2$ there exists a constant $c_p > 0$ independent of N and T such that*

$$\sup_{T \geq 0} \mathbb{E}\left[\left(\mu_T^N(f) - \mu_T(f) \right)^p \right]^{1/p} \leq \frac{c_p \|f\|}{N^{1/2}} \ , \tag{3.4}$$

for any $f \in \mathcal{C}_b(E)$. Furthermore, there exists a constant $c' > 0$ independent of N and T such that

$$\sup_{T \geq 0} \left| \mathbb{E}\left[\mu_T^N(f) \right] - \mu_T(f) \right| \leq \frac{c' \|f\|}{N} \ , \tag{3.5}$$

for any $f \in \mathcal{C}_b(E)$ and $N \in \mathbb{N}$ large enough.

Remark. The constants c_p and c' depend on the Feynman-Kac model of interest, on

the choice of the McKean model and on the considered interacting particle approximation.

The proof, presented in Section 3.2, is an adaptation of the results in Rousset (2006) and makes use of the propagator (2.17) of μ_t and the martingale characterisation of $(\xi_t : t \geq 0)$.

Remark. Theorem 3.1.2 implies in particular that the usual random error $\left|\mu_T^N(f) - \mathbb{E}\left[\mu_T^N(f)\right]\right|$ converges in L^p norm as N goes to infinity, with rate $1/\sqrt{N}$. Moreover, by Markov's inequality, Theorem 3.1.2 implies

$$\mathbb{P}_{\mu_0}\Big(\big|\mu_t^N(f) - \mu_t(f)\big| \geq \varepsilon\Big) \;\leq\; \frac{c_p \cdot \|f\|^p}{\varepsilon^p \cdot N^{p/2}} \;,$$

for every $\varepsilon, t > 0$, $f \in \mathcal{C}_b(E)$, $N \geq K$ and $p \geq 2$, where $c_p > 0$ does not depend on N. In particular, considering $p > 2$, we can see that

$$\mu_t^N(f) \;\to\; \mu_t(f) \qquad \mathbb{P}_{\mu_0} - \text{a.s.} \tag{3.6}$$

as $N \to \infty$, for any $f \in \mathcal{C}_b(E)$, by a Borel-Cantelli argument. The existence of a countable determining class allows this to be further strengthened to the almost sure convergence of μ_t^N to μ_t in the weak topology (see, for example, Schmon et al., 2020, Theorem 4).

Theorem 3.1.3 (Central Limit Theorem)**.** *Consider a sequence of particle approximations satisfying Assumption 3.1.1 with empirical distributions μ_t^N (3.2) and such that*

$$\mu_t^N\big(G_{\mu_t^N}(f,\, f)\big) \;\to\; \mu_t\big(G_{\mu_t}(f,\, f)\big) \qquad \mathbb{P}_{\mu_0} - \text{a.s.} \tag{3.7}$$

as $N \to \infty$, for any $f \in \mathcal{C}_b(E)$ and $0 \leq t \leq T$. Under Assumption 2.2.2 on asymptotic stability, for any $f \in \mathcal{C}_b(E)$, the sequence

$$V_T^N(f) := \sqrt{N}\big(\mu_T^N(f) - \mu_T(f)\big)$$

converges in law as $N \to \infty$ to a centered Gaussian random variable $V_T(f)$ with variance given in terms of the propagator (2.17) as

$$\mathbb{E}\big[V_T(f)^2\big] = \mu_0\big((\Theta_{0,T}\overline{f})^2\big) \;+\; \int_0^T \mu_s\Big(G_{\mu_s}\big(\Theta_{s,T}\overline{f},\; \Theta_{s,T}\overline{f}\big)\Big)\, ds \;, \tag{3.8}$$

where $\overline{f} := f - \mu_T(f)$.

The proof of Theorem 3.1.3 is an adaptation of the results presented by

Del Moral and Miclo (2000, Section 3.3.2) and is outlined in Section 3.3, where we also provide sufficient conditions for the limit (3.7) to hold.

It is important to clarify that the estimators of the Feynman-Kac distribution μ_t given by the empirical measures μ_t^N usually have a bias, i.e. $\mathbb{E}[\mu_t^N(f)] \neq \mu_t(f)$ for $f \in \mathcal{C}_b(E)$, which vanishes only asymptotically, as illustrated in Theorem 3.1.2. This arises from the non-linear time evolution of μ_t. However, it is straightforward to derive unbiased estimators of the unnormalised measures ν_t (2.9), as shown by the following result.

Proposition 3.1.4 (Unbiased Estimators)**.** *Consider a sequence of particle approximations satisfying* (3.3a) *and* (3.3d)*, with empirical distributions* μ_t^N (3.2)*. Then, the unnormalised empirical measure*

$$\nu_t^N(f) := \nu_t^N(1)\mu_t^N(f) \quad \textit{with} \quad \nu_t^N(1) := \exp\Big(\int_0^t \mu_s^N(\mathcal{V})ds\Big) \ ,$$

is an unbiased estimator of the unnormalised t-marginal ν_t (2.9)*, i.e.*

$$\mathbb{E}\big[\nu_t^N(f)\big] = \nu_t(f) \quad \textit{for all } t \geq 0 \textit{ and } N \geq 1 \ , \tag{3.9}$$

for any $f \in \mathcal{C}_b(E)$.

Proof. First observe that $\mathbb{E}\big[\nu_0^N(f)\big] = \nu_0(f)$. Indeed, $\nu_0^N(f) = \mu_0^N(f)$ is the average of N i.i.d. random variables with law $f_{\#}\mu_0$, and μ_0 corresponds to the initial distribution of $\nu_t = \nu_{t,\mu_0}$ (2.9). Here $f_{\#}\mu_0$ denotes the image measure given by $f_{\#}\mu_0(B) = \mu_0(f^{-1}(B))$, for every Borel set B on $\mathbb{R}$.

By Hille-Yosida (Theorem 2.1.5),

$$\frac{d}{dt}\mathbb{E}_{\underline{x}}\big[F(\xi_t)\big] \ = \ \mathbb{E}_{\underline{x}}\big[\overline{L}^N(F)(\xi_t)\big] \ ,$$

for any $F \in \mathcal{C}_b(E^N)$, where ξ_t is the realisation of the interacting particle system at time t. Thus, considering functions of the form $F(\underline{x}) = m(\underline{x})(f)$, $f \in \mathcal{C}_b(E)$, and recalling the definition of the empirical measure μ_t^N (3.2), we can write

$$\frac{d}{dt}\mathbb{E}_{\underline{x}}\big[\mu_t^N(f)\big] \ = \ \mathbb{E}_{\underline{x}}\big[\overline{L}^N(\mu_t^N(f))\big] \ .$$

Hence, by the chain rule, $\mathbb{E}\big[\nu_t^N(f)\big]$ satisfies the evolution equation

$$\frac{d}{dt}\mathbb{E}\big[\nu_t^N(f)\big] = \mathbb{E}\Big[\nu_t^N(f)\,\mu_t^N(\mathcal{V}) + \nu_t^N(1)\,\overline{L}^N\big(\mu_t^N(f)\big)\Big] \ . \tag{3.10}$$

Moreover, by assumption (3.3a) and using the characterisation of $\overline{\mathcal{L}}_\mu$ (2.15)-(2.16), we have

$$\overline{L}^N\big(\mu_t^N(f)\big) = \mu_t^N(\mathcal{L}(f)) + \mu_t^N(\mathcal{V}\, f) - \mu_t^N(\mathcal{V}) \cdot \mu_t^N(f)\ .$$

Inserting into (3.10), this simplifies to

$$\frac{d}{dt}\mathbb{E}\big[\nu_t^N(f)\big] = \mathbb{E}\big[\nu_t^N(\mathcal{L}(f)) + \nu_t^N(\mathcal{V}f)\big]\ .$$

Since $\mathcal{L} + \mathcal{V}$ also generates the time evolution of $\nu_t(f)$ (2.10), a simple Gronwall argument with $\mathbb{E}\big[\nu_0^N(f)\big] = \nu_0(f)$ gives (3.9).

□

3.2 Weak propagation of chaos

This section is devoted to the proof of Theorem 3.1.2, which is an adaptation of the results presented by M. Rousset in Rousset (2006). Throughout this section we consider a generic sequence of IPS generators $(\overline{L}^N)_{N\in\mathbb{N}}$ satisfying Assumption 3.1.1 for some McKean generator $\overline{\mathcal{L}}_\mu$ (2.16). Furthermore, we assume that the normalised Feynman-Kac measure μ_t is asymptotically stable, i.e. Assumption 2.2.2 holds.

The proof makes use of the propagator $\Theta_{t,T}$ of μ_t defined in (2.17), and the martingale characterisation of $\overline{L}^N$. We denote by $\mathcal{C}_b^{0,1}(E\times\mathbb{R}^+)$ the set of bounded functions $\varphi_\cdot$ such that φ_t is continuous on E for every $t\in\mathbb{R}^+$ and $\varphi_\cdot(x)$ has continuous time derivative for every $x\in E$. Following the standard martingale characterisation of Feller-type Markov processes (Lemma 2.1.7), one can show that, for every $\varphi_\cdot \in \mathcal{C}_b^{0,1}(E\times\mathbb{R}^+)$, the process

$$\mathcal{M}_t^N(\varphi_\cdot) \;=\; \mu_t^N(\varphi_t) - \mu_0^N(\varphi_0) - \int_0^t \mu_s^N\big(\partial_s\varphi_s + \overline{\mathcal{L}}_{\mu_s^N}(\varphi_s)\big)\, ds \tag{3.11}$$

is a martingale (see also Rousset, 2006, Proposition 3.3). With (3.3b) its predictable quadratic variation is given by

$$\big\langle \mathcal{M}^N(\varphi_\cdot)\big\rangle_t \;=\; \frac{1}{N}\int_0^t \mu_s^N\big(G_{\mu_s^N}(\varphi_s,\,\varphi_s)\big)\, ds \;+\; O\!\left(\frac{1}{N^2}\right)\ , \tag{3.12}$$

and with (3.3c) jumps are bounded by

$$\big|\Delta\mathcal{M}_t^N(\varphi_\cdot)\big| \;\le\; \frac{2K\,\|\varphi_t\|}{N}\ . \tag{3.13}$$

The following technical Lemma for martingales will play a central role in the

proof of Theorem 3.1.2.

Lemma 3.2.1. *Let $\mathcal{M}$ be a locally square-integrable martingale with predictable quadratic variation $\langle \mathcal{M} \rangle$, $\mathcal{M}_0 = 0$ and uniformly bounded jumps $\sup_t |\Delta \mathcal{M}_t| \leq a < \infty$. Then, for every $q \in \mathbb{N}_0$ and $T \geq 0$, there exists a constant $C_q > 0$ such that*

$$\sup_{t \leq T} \mathbb{E}\left[\mathcal{M}_t^{2^{q+1}}\right] \leq C_q \sum_{k=0}^{q} a^{2^{q+1}-2^{k+1}} \mathbb{E}\left[\left(\langle \mathcal{M} \rangle_T\right)^{2^k}\right].$$

Proof. See (Rousset, 2006, Lemma 6.2). □

3.2.1 Properties of the normalised propagator

Lemma 3.2.2. *For any test function $f \in \mathcal{C}_b(E)$ and $0 \leq t \leq T$, we have for the normalised propagator (2.17)*

$$\partial_t \left(\Theta_{t,T} f(x)\right) = -\big(\mathcal{L} + \mathcal{V}(x) - \mu_t(\mathcal{V})\big)\big(\Theta_{t,T} f(x)\big).$$

Proof. See (Rousset, 2006, p. 836). The idea of the proof is to use the backward equation (2.7) to compute the time derivative of the propagator $\Theta_{t,T} f$ (2.17). □

Lemma 3.2.3. *Under Assumption 2.2.2 on asymptotic stability, for any $0 \leq t \leq T$ and $n \in \mathbb{N}$ and $f \in \mathcal{C}_b(E)$, there exists a constant $\beta > 0$ such that*

$$\|\Theta_{t,T} f\| \leq \beta \cdot \|f\| \qquad \text{and} \qquad \int_t^T \|\Theta_{s,T} f\|^{2^n} ds \leq \beta^{2^n} \cdot \|f\|^{2^n} \cdot (T - t).$$

Moreover, for any $\overline{f} := f - \mu_T(f)$, there exists some $0 < \rho < 1$, such that

$$\|\Theta_{t,T} \overline{f}\| \leq \beta \cdot \|\overline{f}\| \cdot \rho^{T-t} \qquad \text{and} \qquad \int_t^T \|\Theta_{s,T} \overline{f}\|^{2^n} ds \leq \beta^{2^n} \cdot \|\overline{f}\|^{2^n}.$$

Proof. The proof can be found in (Rousset, 2006, Lemma 5.1) and the result is due to the asymptotic stability of the Feynman-Kac model. □

Observe that, applying Lemma 3.2.2 to the martingale characterisation (3.11)

of $\overline{L}^N$, we obtain

$$\begin{aligned}\mathcal{M}_T^N(\Theta_{\cdot,T}f) &= \mu_T^N(f) - \mu_0^N(\Theta_{0,T}f) \; - \int_0^T \mu_s^N\Big(\big(\widetilde{\mathcal{L}}_{\mu_s^N} - \mathcal{V} + \mu_s(\mathcal{V})\big)(\Theta_{s,T}f)\Big)\,ds \\ &= \mu_T^N(f) - \mu_0^N(\Theta_{0,T}f) \; - \int_0^T \mu_s^N(\Theta_{s,T}f)\cdot\big(\mu_s(\mathcal{V}) - \mu_s^N(\mathcal{V})\big)\,ds\;,\end{aligned} \tag{3.14}$$

for any $f \in \mathcal{C}_b(E)$, where the last equality follows by the characterisation (2.18) of McKean models.

By (3.14), we obtain the stochastic differential equation

$$d\mu_t^N(\Theta_{t,T}f) = d\mathcal{M}_t^N(\Theta_{\cdot,T}f) + \big(\mu_t(\mathcal{V}) - \mu_t^N(\mathcal{V})\big)\cdot\mu_t^N(\Theta_{t,T}f)\,dt\;. \tag{3.15}$$

Moreover, applying Lemma 3.2.3 to the predictable quadratic variation (3.12), we obtain that almost surely,

$$\big\langle\mathcal{M}^N(\Theta_{\cdot,T}f)\big\rangle_t \;\leq\; \frac{1}{N}\,\|\overline{G}\|\cdot\beta^2\,\|f\|^2\,(T-t) \;+\; O\Big(\frac{1}{N^2}\Big)\;, \tag{3.16}$$

where $\overline{G}(f,f) = \sup_{\mu\in\mathcal{P}(E)} G_\mu(f,f)$.

Note that Equation (3.14) for centered test functions $\overline{f} = f - \mu_T(f)$ can be rewritten as

$$\mu_T^N(f) - \mu_T(f) \;=\; \mu_0^N(\Theta_{0,T}\overline{f}) + \mathcal{M}_T^N(\Theta_{\cdot,T}\overline{f}) + \int_0^T \mu_s^N\big(\Theta_{s,T}\overline{f}\big)\cdot\big(\mu_s(\mathcal{V}) - \mu_s^N(\mathcal{V})\big)\,ds\;. \tag{3.17}$$

The martingale characterisation (3.14)-(3.17) will be the key element in the proof of Theorem 3.1.2.

3.2.2 L^p and bias estimates

Define

$$\Phi_{t,T}(\mu) := \frac{\mu P^{\mathcal{V}}(T-t)}{\mu\big(P^{\mathcal{V}}(T-t)1\big)} \in \mathcal{P}(E), \tag{3.18}$$

with $\mu \in \mathcal{P}(E)$ and $0 \leq t \leq T$. Observe that the measure $\Phi_{t,T}(\mu)$ can be also rewritten in terms of $\Theta_{t,T}$ (2.17) as

$$\Phi_{t,T}(\mu)(f) \;=\; \frac{\mu(\Theta_{t,T}f)}{\mu(\Theta_{t,T}1)}\;, \tag{3.19}$$

for any $f \in \mathcal{C}_b(E)$.

To prove Theorem 3.1.2, we consider the decomposition

$$\mathbb{E}[|\mu_T^N(f) - \mu_T(f)|^p]^{1/p} \leq \mathbb{E}[|\mu_T^N(f) - \Phi_{t,T}(\mu_t^N)(f)|^p]^{1/p} + \mathbb{E}[|\Phi_{t,T}(\mu_t^N)(f) - \mu_T(f)|^p]^{1/p}, \tag{3.20}$$

for any $0 \leq t \leq T$. The proof is structured as follows:

- In Lemma 3.2.4, we bound the first term of the decomposition under Assumptions 2.2.2 and 3.1.1;
- In Lemma 3.2.5, we bound the second term under Assumption 2.2.2;
- In Lemma 3.2.6, we combine Lemma 3.2.4 and Lemma 3.2.5 to obtain L^p-error estimates of order $1/N^{\delta/2}$, for some $\delta \in (0, 1)$;
- Finally, from Lemma 3.2.6 we derive, by iteration, L^p estimates of order $1/\sqrt{N}$, as presented in Theorem 3.1.2.

Lemma 3.2.4. *Consider a sequence of particle approximations satisfying Assumption 3.1.1 with empirical distributions μ_t^N (3.2). Under Assumption 2.2.2 on asymptotic stability, for any $p \geq 2$ there exists a constant $c_p > 0$ such that*

$$\mathbb{E}\Big[\big|\mu_T^N(f) - \Phi_{t,T}(\mu_t^N)(f)\big|^p\Big] \leq c_p\, e^{4p(T-t)\|\mathcal{V}\|} \left(\frac{\|f\|^p\,(T-t)^{p/2}}{N^{p/2}} + O\Big(\frac{1}{N^p}\Big) \right),$$

for any $f \in \mathcal{C}_b(E)$ and $0 \leq t \leq T$.

Proof. This is an adaptation of the first part of the proof of (Rousset, 2006, Lemma 5.3).

First, consider

$$A_{t_1}^{t_2} := \exp\Big(\int_{t_1}^{t_2} \big(\mu_s^N(\mathcal{V}) - \mu_s(\mathcal{V})\big)\, ds \Big), \tag{3.21}$$

with $0 \leq t_1 \leq t_2$. Observe that, by the stochastic differential equation (3.15), we can write

$$d\big(A_t^s \mu_s^N(\Theta_{s,T} f)\big) = A_t^s\, d\mathcal{M}_s^N(\Theta_{\cdot,T} f),$$

for any $t \leq s \leq T$. Therefore,

$$A_t^T \mu_T^N(f) - \mu_t^N(\Theta_{t,T} f) = \int_t^T A_t^s\, d\mathcal{M}_s^N(\Theta_{\cdot,T} f). \tag{3.22}$$

Fixing $0 \leq t \leq T$, the process

$$\mathcal{N}_\tau^N(f) := \int_t^\tau A_t^s \, d\mathcal{M}_s^N(\Theta_{\cdot,T} f) \;=\; A_t^\tau \cdot \mu_\tau^N\big(\Theta_{\tau,T} f\big) \;-\; \mu_t^N(\Theta_{t,T} f) \;,$$

with $t \leq \tau \leq T$, as the integral of a progressively measurable process with respect to a martingale, is itself a martingale with predictable quadratic variation given by

$$\langle \mathcal{N}^N(f) \rangle_\tau = \int_t^\tau \big(A_t^s\big)^2 d\langle \mathcal{M}_s^N(\Theta_{\cdot,T} f) \rangle \;,$$

and jumps bounded by

$$\big|\Delta \mathcal{N}_\tau^N(f)\big| \;\leq\; e^{2(T-t)\|\mathcal{V}\|} \cdot \frac{4K\,\beta\,\|f\|}{N} \;,$$

by Assumption (3.3c) on bounded jumps, (3.13) and Lemma 3.2.3.

Moreover, with (3.19), we can write

$$\begin{aligned}
&\big|\mu_T^N(f) - \Phi_{t,T}(\mu_t^N)(f)\big| \\
&\quad = \Big| \mu_T^N(f) - (A_t^T)^{-1} \mu_t^N(\Theta_{t,T} f) - \Big(1 - (A_t^T)^{-1} \mu_t^N(\Theta_{t,T} 1)\Big) \cdot \Phi_{t,T}(\mu_t^N)(f) \Big| \\
&\quad = (A_t^T)^{-1} \Big| \mathcal{N}_T^N(f) \;-\; \mathcal{N}_T^N(1) \cdot \Phi_{t,T}(\mu_t^N)(f) \Big| \;,
\end{aligned}$$

where the last equality follows by (3.22). Noting that $(A_t^T)^{-1} \leq \exp\big(2(T-t) \cdot \|\mathcal{V}\|\big)$ by definition (3.21), we get

$$\begin{aligned}
&\mathbb{E}\Big[\big|\mu_T^N(f) - \Phi_{t,T}(\mu_t^N)(f)\big|^p\Big] \\
&\qquad\qquad \leq\; e^{2p(T-t)\|\mathcal{V}\|}\, \mathbb{E}\Big[\, \Big| \mathcal{N}_T^N(f) \;-\; \mathcal{N}_T^N(1) \cdot \Phi_{t,T}(\mu_t^N)(f) \Big|^p \Big] \;. \qquad (3.23)
\end{aligned}$$

By Lemma 3.2.1, we have that, for any $q \in \mathbb{N}_0$,

$$\mathbb{E}\Big[\big|\mathcal{N}_T^N(f)\big|^{2^{q+1}}\Big] \;\leq C_q \sum_{k=0}^{q} \Big(e^{2(T-t)\|\mathcal{V}\|} \cdot \frac{2K\,\beta\|f\|}{N}\Big)^{2^{q+1}-2^{k+1}} \mathbb{E}\Big[\big(\langle \mathcal{N}_\cdot^N(f) \rangle_T\big)^{2^k}\Big] \;,$$

and thus, by (3.16),

$$\mathbb{E}\Big[\big|\mathcal{N}_T^N(f)\big|^{2^{q+1}}\Big] \;\leq \widetilde{C}_q \sum_{k=0}^{q} \Big(e^{2(T-t)\|\mathcal{V}\|} . \frac{\|f\|}{N}\Big)^{2^{q+1}-2^{k+1}} \Big(\frac{1}{N}\,\|f\|^2\,(T-t) + O\Big(\frac{1}{N^2}\Big)\Big)^{2^k} \;.$$

Therefore, for $p = 2^{q+1}$, $q \in \mathbb{N}_0$, we get

$$\mathbb{E}\Big[\big|\mathcal{N}_T^N(f)\big|^p\Big] \leq \widetilde{C}_p\, e^{2p(T-t)\|\mathcal{V}\|} \left(\frac{\|f\|^p\,(T-t)^{p/2}}{N^{p/2}} + O\Big(\frac{1}{N^p}\Big)\right).$$

By Jensen's inequality, this bound holds for any $p \geq 2$. Applying this to inequality (3.23), we obtain the result.

□

Lemma 3.2.5. *Under Assumption 2.2.2 on asymptotic stability with constants $\alpha > 0$ and $\rho \in (0,1)$, we have that for any $p \geq 2$ and any $0 \leq t \leq T$ such that $T - t \geq (\log\varepsilon - \log\alpha)/\log\rho$ for some $\varepsilon \in (0,1)$, the following bound holds*

$$\mathbb{E}\big[|\Phi_{t,T}(\mu_t^N)(f) - \mu_T(f)|^p\big]^{1/p} \leq \frac{4\|f\|\,\alpha\rho^{T-t}}{1-\varepsilon}.$$

Furthermore, when $t = 0$, there exists a constant $C_p > 0$ depending on p such that

$$\sup_{T\geq 0} \mathbb{E}\big[\big|\Phi_{0,T}(\mu_0^N)(f) - \mu_T(f)\big|^p\big]^{1/p} \leq \frac{C_p\,\|f\|}{N^{1/2}}.$$

Proof. By definition (3.18) of $\Phi_{t,T}$, for any $\eta \in \mathcal{P}(E)$ and $\lambda \in \mathbb{R}$ we have

$$\Phi_{t,T}(\eta)(f) = \frac{\eta\big(e^{-(T-t)\lambda}P^{\mathcal{V}}(T-t)f\big)}{\eta\big(e^{-(T-t)\lambda}P^{\mathcal{V}}(T-t)1\big)}.$$

Taking λ to be the principal eigenvalue of $\mathcal{L}+\mathcal{V}$, using Assumption 2.2.2 on asymptotic stability and the basic fact $\eta(1) = 1$, we can write

$$\begin{aligned}\eta\big(e^{-(T-t)\lambda}P^{\mathcal{V}}(T-t)f\big) &\leq \mu_\infty(f) + \|f\|\cdot\alpha\rho^{T-t},\\ \eta\big(e^{-(T-t)\lambda}P^{\mathcal{V}}(T-t)1\big) &\geq 1 - \alpha\rho^{T-t}.\end{aligned}$$

Therefore, for $T - t \geq (\log\varepsilon - \log\alpha)/\log\rho$, for some $\varepsilon \in (0,1)$, we have

$$\begin{aligned}\Phi_{t,T}(\eta)(f) - \mu_\infty(f) &\leq \mu_\infty(f)\cdot\Big(\frac{1}{1-\alpha\rho^{T-t}} - 1\Big) + \frac{\|f\|\alpha\rho^{T-t}}{1-\alpha\rho^{T-t}}\\ &\leq \frac{2\|f\|\,\alpha\rho^{T-t}}{1-\varepsilon},\end{aligned}$$

and similarly

$$\Phi_{t,T}(\eta)(f) - \mu_\infty(f) \geq -\frac{2\|f\|\,\alpha\rho^{T-t}}{1-\varepsilon}.$$

Hence,

$$\begin{aligned}
&\mathbb{E}\big[|\Phi_{t,T}(\mu_t^N)(f) - \mu_T(f)|^p\big]^{1/p} \\
&\quad \le \mathbb{E}\big[|\Phi_{t,T}(\mu_t^N)(f) - \mu_\infty(f)|^p\big]^{1/p} + \mathbb{E}\big[|\Phi_{t,T}(\mu_t)(f) - \mu_\infty(f)|^p\big]^{1/p} \\
&\quad \le \frac{4\|f\|\,\alpha\rho^{T-t}}{1-\varepsilon} \ .
\end{aligned}$$

This proves the first part of Lemma 3.2.5. Now, for $t = 0$, by definition of $\Phi_{0,T}$, we have

$$\Phi_{0,T}(\mu_0^N)(f) \ = \ \frac{\mu_0^N\big(\Theta_{0,T}(f)\big)}{\mu_0^N\big(\Theta_{0,T}(1)\big)} \ .$$

Thus, we can write

$$\begin{aligned}
&\Phi_{0,T}(\mu_0^N)(f) \ - \ \mu_T(f) \\
&\quad = \ \Phi_{0,T}(\mu_0^N)(f) \cdot \big(1 - \mu_0^N(\Theta_{0,T}(1))\big) \ + \ \mu_0^N\big(\Theta_{0,T}(f)\big) - \mu_0\big(\Theta_{0,T}(f)\big) \\
&\quad = \ \Phi_{0,T}(\mu_0^N)(f) \cdot \big(\mu_0(\Theta_{0,T}(1)) - \mu_0^N(\Theta_{0,T}(1))\big) \ + \ \mu_0^N\big(\Theta_{0,T}(f)\big) - \mu_0\big(\Theta_{0,T}(f)\big) \ ,
\end{aligned}$$

by using the basic substitution $1 = \mu_0(\Theta_{0,T}(1)$. To conclude it is enough to observe that, for any $g \in \mathcal{C}_b(E)$,

$$\mathbb{E}\big[\big|\mu_0^N(g) - \mu_0(g)\big|^p\big] \le \frac{C_p\,\|g\|^p}{N^{p/2}}, \tag{3.24}$$

with $C_p > 0$ constant depending on p (so, we can apply (3.24) for $g = \Theta_{0,T}(1)$ and $g = \Theta_{0,T}(f)$). Indeed, with (3.3d) at time $t = 0$, μ_0^N is the sum of N i.i.d. random variables with law μ_0. Inequality (3.24) is then a direct application of Marcinkiewicz-Zygmund/BDG inequalities for i.i.d. variables.

□

Lemma 3.2.6. *Consider a sequence of particle approximations satisfying Assumption 3.1.1 with empirical distributions μ_t^N (3.2). Under Assumption 2.2.2, there exists $\delta \in (0,1)$ such that for any $p \ge 2$ there exists $c_p > 0$ such that*

$$\sup_{T\ge 0} \mathbb{E}[|\mu_T^N(f) - \mu_T(f)|^p]^{1/p} \ \le \ \frac{c_p\,\|f\|}{N^{\delta/2}} \ ,$$

for any $N \in \mathbb{N}$ large enough.

Proof. Recalling decomposition (3.20), where the first term is estimated in Lemma 3.2.4 and the second in Lemma 3.2.5, and using the basic fact $T - t \le e^{T-t}$, we

obtain

$$\mathbb{E}[|\mu_T^N(f) - \mu_T(f)|^p]^{1/p} \\ \leq c_p \, \|f\| \cdot \frac{e^{(4\|\mathcal{V}\|+1/2)T} + 1}{N^{1/2}} + e^{4T\|\mathcal{V}\|} \, O\Big(\frac{1}{N}\Big) \, , \tag{3.25}$$

taking $t = 0$, and

$$\mathbb{E}[|\mu_T^N(f) - \mu_T(f)|^p]^{1/p} \\ \leq c_p \|f\| \cdot \Big(\frac{e^{(4\|\mathcal{V}\|+1/2)\cdot(T-t)}}{N^{1/2}} + \rho^{T-t} \Big) + e^{4(T-t)\|\mathcal{V}\|} \, O\Big(\frac{1}{N}\Big) \, , \tag{3.26}$$

taking $0 \leq t \leq T$ such that $T - t$ is large enough.

The idea is to find $t \geq 0$ and $\varepsilon \in (0, 1)$ such that

$$\begin{cases} \frac{e^{(4\|\mathcal{V}\|+1/2)\cdot(T-t)}}{N^{1/2}} \leq \frac{1}{N^{\varepsilon/2}} \, , \\ \rho^{T-t} \leq \frac{1}{N^{\varepsilon/2}} \, . \end{cases}$$

Recalling that $\log \rho < 0$, the solution is given by

$$\begin{cases} \varepsilon \, = \, \frac{-\log \rho}{4\|\mathcal{V}\| + \frac{1}{2} - \log \rho} \, , \\ t \, = \, T - \frac{\log N}{8\|\mathcal{V}\| + 1 - 2\log \rho} \, , \end{cases} \tag{3.27}$$

provided $T \geq \log N/(8\|\mathcal{V}\| + 1 - 2\log \rho)$ to ensure that $t \geq 0$. Also observe that for N large enough, $T - t$ satisfies the conditions in Lemma 3.2.5.

Otherwise, in case $T < \log N/(8\|\mathcal{V}\| + 1 - 2\log \rho)$, we consider the bound (3.25) instead, and we obtain

$$\frac{e^{(4\|\mathcal{V}\|+1/2)T} + 1}{N^{1/2}} \leq \frac{1}{N^{\bar{\varepsilon}/2}} + \frac{1}{N^{1/2}} \, ,$$

with

$$\bar{\varepsilon} \, = \, 1 \, - \, \frac{8\|\mathcal{V}\| + 1}{8\|\mathcal{V}\| + 1 - 2\log \rho} \, .$$

Taking $\delta = \min\{\varepsilon, \bar{\varepsilon}\}$ the result follows from observing that

$$\frac{e^{4(T-t)\|\mathcal{V}\|}}{N} \, = \, \frac{1}{N^{\alpha}} \, , \quad \text{with } \alpha > \frac{1}{2} \, ,$$

for $t = 0$ and T at most of order $\log N$ as above, or for $t \geq 0$ given by (3.27).

□

Proof of Theorem 3.1.2. We denote

$$I_p(N) := \sup_{\|g\|=1} \sup_{T\geq 0} \mathbb{E}\big[\big|\mu_T^N(g) - \mu_T(g)\big|^p\big] ,$$

in accordance with Rousset (2006, Section 5.2).

Using (3.17), we have for any $\overline{f} = f - \mu_T(f)$, $f \in \mathcal{C}_b(E)$,

$$\begin{aligned}\big|\mu_T^N(f) - \mu_T(f)\big|^p \leq\ & 3^p \big|\mu_0^N(\Theta_{0,T}\overline{f})\big|^p + 3^p \big|\mathcal{M}_T^N(\Theta_{\cdot,T}\overline{f})\big|^p + \\ & 3^p \Big(\int_0^T \big|\mu_s^N\big(\Theta_{s,T}\overline{f}\big)\big| \cdot \big|\mu_s^N(\mathcal{V}) - \mu_s(\mathcal{V})\big|\, ds\Big)^p . \qquad (3.28)\end{aligned}$$

Writing

$$\begin{aligned}&\big|\mu_s^N\big(\Theta_{s,T}\overline{f}\big)\big| \cdot \big|\mu_s^N(\mathcal{V}) - \mu_s(\mathcal{V})\big| \\ &\quad = \|\Theta_{s,T}\overline{f}\|^{1-1/p} \cdot \Big(\Big|\mu_s^N\Big(\frac{\Theta_{s,T}\overline{f}}{\|\Theta_{s,T}\overline{f}\|}\Big)\Big| \cdot \|\Theta_{s,T}\overline{f}\|^{1/p} \cdot \big|\mu_s^N(\mathcal{V}) - \mu_s(\mathcal{V})\big|\Big) ,\end{aligned}$$

and using Hölder's inequality, we get that the third term in (3.28) can be bounded by

$$\begin{aligned}&\Big(\int_0^T \big|\mu_s^N\big(\Theta_{s,T}\overline{f}\big)\big| \cdot \big|\mu_s^N(\mathcal{V}) - \mu_s(\mathcal{V})\big|\, ds\Big)^p \\ &\leq \Big(\int_0^T \|\Theta_{s,T}\overline{f}\|\, ds\Big)^{p-1} \cdot \Big(\int_0^T \Big|\mu_s^N\Big(\frac{\Theta_{s,T}\overline{f}}{\|\Theta_{s,T}\overline{f}\|}\Big)\Big|^p \cdot \|\Theta_{s,T}\overline{f}\| \cdot \big|\mu_s^N(\mathcal{V}) - \mu_s(\mathcal{V})\big|^p\, ds\Big) \\ &\leq C_p\|\overline{f}\|^{p-1} \Big(\int_0^T \Big|\mu_s^N\Big(\frac{\Theta_{s,T}\overline{f}}{\|\Theta_{s,T}\overline{f}\|}\Big)\Big|^p \cdot \|\Theta_{s,T}\overline{f}\| \cdot \big|\mu_s^N(\mathcal{V}) - \mu_s(\mathcal{V})\big|^p\, ds\Big) ,\end{aligned}$$

by Lemma 3.2.3. Using the fact that

$$\mu_s^N\big(\Theta_{s,T}\overline{f}\big) = \mu_s^N\big(\Theta_{s,T}\overline{f}\big) - \mu_s\big(\Theta_{s,T}\overline{f}\big) ,$$

for centered test functions, and applying the Cauchy-Schwarz inequality, we get

$$\begin{aligned}
&\mathbb{E}\Big[\int_0^T \Big|\mu_s^N\Big(\frac{\Theta_{s,T}\overline{f}}{\|\Theta_{s,T}\overline{f}\|}\Big)\Big|^p \cdot \big|\mu_s^N(\mathcal{V}) - \mu_s(\mathcal{V})\big|^p \cdot \|\Theta_{s,T}\overline{f}\|\, ds\Big] \\
&\qquad \le \int_0^T \mathbb{E}\Big[\Big|\mu_s^N\Big(\frac{\Theta_{s,T}\overline{f}}{\|\Theta_{s,T}\overline{f}\|}\Big) - \mu_s\Big(\frac{\Theta_{s,T}\overline{f}}{\|\Theta_{s,T}\overline{f}\|}\Big)\Big|^{2p}\Big]^{1/2} \\
&\qquad\quad \cdot \|\mathcal{V}\|^p\, \mathbb{E}\Big[\Big|\mu_s^N\Big(\frac{\mathcal{V}}{\|\mathcal{V}\|}\Big) - \mu_s\Big(\frac{\mathcal{V}}{\|\mathcal{V}\|}\Big)\Big|^{2p}\Big]^{1/2} \cdot \|\Theta_{s,T}\overline{f}\|\, ds \\
&\qquad \le \int_0^T I_{2p}(N)\, \|\mathcal{V}\|^p \cdot \|\Theta_{s,T}\overline{f}\|\, ds \\
&\qquad \le C_p \|f\|\, I_{2p}(N)\,. \qquad (3.29)
\end{aligned}$$

Furthermore, similarly to (3.24), we can bound the first term in (3.28) by

$$\mathbb{E}\big[\big|\mu_0^N(\Theta_{0,T}\overline{f})\big|^p\big] \;=\; \mathbb{E}\big[\big|\mu_0^N(\Theta_{0,T}f) - \mu_0(\Theta_{0,T}f)\big|^p\big] \;\le\; \frac{C_p\|f\|^p}{N^{p/2}}\,,$$

for some other constant $C_p > 0$ depending on p. Finally, by Lemma 3.2.1 and bound (3.16), we can bound the second term in (3.28) by

$$\mathbb{E}\big[\big|\mathcal{M}_T^N(\Theta_{\cdot,T}\overline{f})\big|^p\big] \;\le\; \frac{C_p\|f\|^p}{N^{p/2}}\,,$$

where C_p is another p-dependent constant.

Combining all together, we obtain

$$\mathbb{E}\big[\big|\mu_T^N(f) \,-\, \mu_T(f)\big|^p\big] \;\le C_p\|f\|^p \Big(\frac{1}{N^{p/2}} \,+\, I_{2p}(N)\Big)\,,$$

for any $f \in \mathcal{C}_b(E)$ and $T \ge 0$. In particular,

$$I_p(N) \;\le\; C_p \Big(\frac{1}{N^{p/2}} \,+\, I_{2p}(N)\Big)\,, \qquad (3.30)$$

for any $p \ge 2$. Applying Lemma 3.2.6, we get

$$I_p(N) \;\le\; \frac{C_p}{N^{\min\{1,2^k\delta\}p/2}}\,,$$

for any $k \in \mathbb{N}$, by iteration of (3.30). Thus, we can conclude

$$I_p(N) \;\le\; \frac{C_p}{N^{p/2}}\,.$$

This proves the L^p-error estimate (3.4).

We conclude by proving the bias estimate (3.5). By Equation (3.17), we have

$$\mathbb{E}\big[\mu_T^N(f)\big] \,-\, \mu_T(f) \,=\, \int_0^T \|\Theta_{s,T}\overline{f}\| \cdot \mathbb{E}\Big[\mu_s^N\Big(\frac{\Theta_{s,T}\overline{f}}{\|\Theta_{s,T}\overline{f}\|}\Big) \cdot \big(\mu_s(\mathcal{V}) \,-\, \mu_s^N(\mathcal{V})\big)\Big]\, ds \;.$$

Using (3.29) for $p = 1$, we obtain

$$\big|\mathbb{E}\big[\mu_T^N(f)\big] \,-\, \mu_T(f)\big| \,\leq\, C\|f\| \cdot I_2(N) \,\leq\, \frac{C\|f\|}{N} \;.$$

□

3.3 Central Limit Theorem

In this section, we prove the central limit theorem (Theorem 3.1.3), following similar arguments as in Del Moral and Miclo (2000), Section 3.3.2, and, similarly to the proof of Theorem 3.1.2, it makes use of the propagator $\Theta_{t,T}$ (2.17).

Recall Equation (3.17), that is

$$\mu_T^N(f) - \mu_T(f) \,=\, \mu_0^N(\Theta_{0,T}\overline{f}) + \mathcal{M}_T^N(\Theta_{\cdot,T}\overline{f}) + \int_0^T \mu_s^N\big(\Theta_{s,T}\overline{f}\big) \cdot \big(\mu_s(\mathcal{V}) - \mu_s^N(\mathcal{V})\big)\, ds \;,$$

where the martingale $\mathcal{M}_T^N\big(\Theta_{\cdot,T}\overline{f}\big)$ has predictable quadratic variation given by

$$\big\langle \mathcal{M}^N\big(\Theta_{\cdot,T}\overline{f}\big)\big\rangle_t \,=\, \frac{1}{N}\int_0^t \mu_s^N\big(G_{\mu_s^N}(\Theta_{s,T}\overline{f},\, \Theta_{s,T}\overline{f})\big)\, ds \,+\, O\Big(\frac{1}{N^2}\Big) \;, \tag{3.31}$$

and jumps bounded by

$$\big|\Delta\mathcal{M}_t^N\big(\Theta_{\cdot,T}\overline{f}\big)\big| \,\leq\, \frac{2K\beta\,\|f\|}{N} \;.$$

If the hypotheses of Theorem 3.1.3 are satisfied, we have that

$$\int_0^T \mu_s^N\big(G_{\mu_s^N}(\Theta_{s,T}\overline{f},\, \Theta_{s,T}\overline{f})\big)\, ds \;\to\; \int_0^T \mu_s\big(G_{\mu_s}(\Theta_{s,T}\overline{f},\, \Theta_{s,T}\overline{f})\big)\, ds \qquad \mathbb{P}_{\mu_0}-\text{a.s.}$$

as $N \to \infty$, for any $\overline{f} = f - \mu_T(f)$ centered test function, by using Theorem 3.1.2 and dominated convergence. In particular, by (3.31), we see that

$$\lim_{N\to\infty} N \cdot \big\langle \mathcal{M}^N(\Theta_{\cdot,T}\overline{f})\big\rangle_T \,=\, \int_0^T \mu_s\big(G_{\mu_s}(\Theta_{s,T}\overline{f},\, \Theta_{s,T}\overline{f})\big)\, ds \;,$$

in probability. Thus, in the same fashion as Del Moral and Miclo (2000, Lemma

3.33), we can apply (Jacod and Shiryaev, 2013, Theorem 3.11, p. 432) and obtain that the process $\sqrt{N}\cdot\mathcal{M}_T^N(\Theta_{\cdot,T}f)$ converges in law as $N\to\infty$ to a centered Gaussian martingale whose variance is given by

$$\int_0^T \mu_s\big(G_{\mu_s}(\Theta_{s,T}\overline{f},\,\Theta_{s,T}\overline{f})\big)\,ds\ .$$

Moreover, the following result shows that

$$\lim_{N\to\infty}\sqrt{N}\int_0^T\Big|\mu_s^N(\Theta_{s,T}\overline{f})\cdot\big(\mu_s(\mathcal{V})-\mu_s^N(\mathcal{V})\big)\Big|\,ds\ =\ 0\ ,$$

in probability.

Lemma 3.3.1. *Consider a sequence of particle approximations satisfying Assumption 3.1.1 with empirical distributions μ_t^N (3.2). If Assumption 2.2.2 on asymptotic stability is satisfied, then*

$$\lim_{N\to\infty}\mathbb{E}\left[\sqrt{N}\int_0^T\Big|\mu_s^N(\Theta_{s,T}\overline{f})\cdot\big(\mu_s(\mathcal{V})-\mu_s^N(\mathcal{V})\big)\Big|\,ds\right]\ =\ 0\ .$$

Proof. This proof is an adaptation of (Del Moral and Miclo, 2000, Lemma 3.30). Using the Cauchy-Schwarz inequality, we can write

$$\begin{aligned}&\mathbb{E}\Big[\sqrt{N}\int_0^T\Big|\mu_s^N(\Theta_{s,T}\overline{f})\cdot\big(\mu_s^N(\mathcal{V})-\mu_s(\mathcal{V})\big)\Big|\,ds\Big]\\&\quad\le\sqrt{N\,\mathbb{E}\Big[\int_0^T\big(\mu_s^N(\Theta_{s,T}\overline{f})\big)^2\,ds\Big]}\cdot\sqrt{\mathbb{E}\Big[\int_0^T\big(\mu_s^N(\mathcal{V})-\mu_s(\mathcal{V})\big)^2\,ds\Big]}.\end{aligned}$$

By Theorem 3.1.2 and dominated convergence, the second factor goes to zero as $N\to\infty$. Moreover, for any $0\le s\le T$, we can write

$$\frac{\mu_s^N(\Theta_{s,T}\overline{f})}{\mu_s^N(\Theta_{s,T}1)}\ =\ \frac{\mu_s^N(\Theta_{s,T}f)}{\mu_s^N(\Theta_{s,T}1)}\ -\ \frac{\mu_s(\Theta_{s,T}f)}{\mu_s(\Theta_{s,T}1)}\ ,$$

by using the basic fact

$$\frac{\mu_s^N\big(\Theta_{s,T}\mu_T(f)\big)}{\mu_s^N(\Theta_{s,T}1)}\ =\ \mu_T(f)\ =\ \frac{\mu_s(\Theta_{s,T}f)}{\mu_s(\Theta_{s,T}1)}\ .$$

Therefore,

$$\Big(\frac{\mu_s^N(\Theta_{s,T}\overline{f})}{\mu_s^N(\Theta_{s,T}1)}\Big)^2 \leq 2\,\frac{\big(\mu_s^N(\Theta_{s,T}f)-\mu_s(\Theta_{s,T}f)\big)^2}{\mu_s^N(\Theta_{s,T}1)^2} + 2\,\frac{\mu_s(\Theta_{s,T}f)^2\cdot\big(\mu_s(\Theta_{s,T}1)-\mu_s^N(\Theta_{s,T}1)\big)^2}{\mu_s^N(\Theta_{s,T}1)^2\mu_s(\Theta_{s,T}1)^2}\ .$$

The conclusion follows by observing that

$$\sup_{0\leq s\leq T} N\,\mathbb{E}\Big[\big(\mu_s^N(\Theta_{s,T}f)-\mu_s(\Theta_{s,T}f)\big)^2\Big] \ < \ \infty\ ,$$

for any $f\in\mathcal{C}_b(E)$, by Theorem 3.1.2.

□

Finally, recalling that $\mu_0^N(\Theta_{0,T}\overline{f})$ converges to $\mu_0(\Theta_{0,T}\overline{f})=0$ in probability and that $\mathbb{E}\big[N\cdot\mu_0^N\big((\Theta_{0,T}\overline{f})^2\big)\big]=\mu_0\big((\Theta_{0,T}\overline{f})^2\big)$, we obtain the statement.

The following result provides sufficient conditions for the sequence $\mu_t^N\big(G_{\mu_t^N}(f,f)\big)$ to satisfy the a.s. convergence (3.7), for any $f\in\mathcal{C}_b(E)$.

Lemma 3.3.2. *Consider a sequence of particle approximations satisfying Assumption 3.1.1 with empirical distributions μ_t^N (3.2) and such that $G_\mu(f,f)$ is of the form*

$$G_\mu(f,\,f)(x) \ = \ g_0(x) \ + \ \int_E g_1(y,x)\,\mu(dy) \ + \ \Big(\int_E g_2(y,x)\,\mu(dy)\Big)^2\ , \tag{3.32}$$

for any $\mu\in\mathcal{P}(E)$, $f\in\mathcal{C}_b(E)$ and $x\in E$, where $g_0\in\mathcal{C}_b(E)$, g_1, $g_2\in\mathcal{C}_b(E^2)$ depend on f but are independent of μ, and such that

$$\lim_{z\to x}\ \|g_i(\cdot,z)-g_i(\cdot,x)\| \ = \ 0\ , \tag{3.33}$$

for any $x\in E$, with $i=1,2$. Then,

$$\mu_t^N\big(G_{\mu_t^N}(f,f)\big) \ \xrightarrow{a.s.}\ \mu_s\big(G_{\mu_t}(f,f)\big) \quad \text{as } N\to\infty\ .$$

Remark. Note that for discrete state spaces, condition (3.33) is always satisfied with the usual topology.

Proof. The statement is a direct application of a variant of Lebesgue's dominated convergence theorem (Feinberg et al., 2020, Corollary 2.8) which applies in the context of weakly converging measures (see also Serfozo, 1982, Theorem 3.5). Indeed, μ_t^N converges a.s. to μ_t in the weak topology via the argument following (3.6). Moreover, $G_{\mu_t^N}(f,\,f)\in\mathcal{C}_b(E)$ and $\sup_{N\in\mathbb{N}}\|G_{\mu_t^N}(f,f)\|<\infty$ by condition (3.3b), hence

the sequence of functions $\big(G_{\mu_t^N}(f,\,f)\big)_{N\in\mathbb{N}}$ can easily be seen to be asymptotically uniformly integrable w.r.t. the empirical measures μ_t^N (see Feinberg et al., 2020, Definition 2.1). Finally, we need to show that

$$\lim_{(z,N)\to(x,\infty)} G_{\mu_t^N}(f,f)(z) \;=\; G_{\mu_t}(f,f)(x) \qquad \mathbb{P}_{\mu_0}-\text{a.s.} \tag{3.34}$$

for any $x\in E$. Indeed, by triangular inequality,

$$\begin{aligned}\Big|\int_E g_1(y,z)\,\mu_t^N(dy) \;-\; \int_E g_1(y,x)\,\mu_t(dy)\Big| \\ \leq \Big|\int_E g_1(y,z)\,\mu_t^N(dy) \;-\; \int_E g_1(y,x)\,\mu_t^N(dy)\Big| \\ +\;\Big|\int_E g_1(y,x)\,\mu_t^N(dy) \;-\; \int_E g_1(y,x)\,\mu_t(dy)\Big|\,,\end{aligned}$$

and

$$\Big|\int_E g_1(y,z)\,\mu_t^N(dy) \;-\; \int_E g_1(y,x)\,\mu_t^N(dy)\Big| \;\leq\; \|g(\cdot,z) \;-\; g(\cdot,x)\|\,.$$

Thus,

$$\lim_{(z,N)\to(x,\infty)} \int_E g_1(y,z)\,\mu_t^N(dy) \;=\; \int_E g_1(y,x)\,\mu_t(dy) \qquad \mathbb{P}_{\mu_0}-\text{a.s.}$$

for any $x\in E$, by condition (3.33) on g_1 and by a.s. weak convergence of μ_t^N. Moreover,

$$\begin{aligned}\Big|\Big(\int_E g_2(y,z)\,\mu_t^N(dy)\Big)^2 \;-\; \Big(\int_E g_2(y,x)\,\mu_t(dy)\Big)^2\Big| \\ \leq\; \Big|\int_E g_2(y,z)\,\mu_t^N(dy) \;+\; \int_E g_2(y,x)\,\mu_t(dy)\Big| \\ \cdot\;\Big|\int_E g_2(y,z)\,\mu_t^N(dy) \;-\; \int_E g_2(y,x)\,\mu_t(dy)\Big| \\ \leq 2\,\|g_2\|\cdot\Big|\int_E g_2(y,z)\,\mu_t^N(dy) \;-\; \int_E g_2(y,x)\,\mu_t(dy)\Big|\,.\end{aligned}$$

Thus, similarly to g_1, we obtain

$$\lim_{(z,N)\to(x,\infty)} \Big(\int_E g_2(y,z)\,\mu_t^N(dy)\Big)^2 \;=\; \Big(\int_E g_2(y,x)\,\mu_t(dy)\Big)^2 \qquad \mathbb{P}_{\mu_0}-\text{a.s.}$$

for any $x\in E$ and, in particular, we see that (3.34) holds. Therefore, the hypotheses in (Feinberg et al., 2020, Corollary 2.8) are satisfied almost surely and we conclude by applying it.

□

Lemma 3.3.2 provides a sufficient condition that is easy to verify. In particular, we will show that (3.32) holds for both mean field approximations and cloning algorithms, as illustrated respectively in Proposition 4.1.1 and Proposition 4.2.5.

Chapter 4

Description of some Interacting Particle Models

In this chapter we describe two common constructions for interacting particle approximations, namely mean field particle approximations (in Section 4.1) and cloning algorithms (in Section 4.2), and provide a discussion on their convergence properties by applying the results presented in Chapter 3. In Section 4.3, we present a third interacting particle approximation - the continuous time resampling algorithm - and show that in this case Assumption 3.1.1 is not satisfied and the convergence results presented in Chapter 3 do not hold, in disagreement with the well-established convergence results for discrete time resampling algorithms.

4.1 Mean field particle approximation

The most basic particle approximation is simply to run the McKean dynamics in parallel on each of the particles, replacing the distribution μ_t by the empirical measure. This procedure has been studied in the applied probability literature in great detail (Del Moral and Miclo, 2003; Rousset, 2006), providing quantitative control on error bounds for convergence.

Formally, the mean field particle model $(\xi_t : t \geq 0)$ with $\xi_t = (\xi_t^i : i = 1, \ldots, N)$ associated to a McKean generator $\overline{\mathcal{L}}_{\mu_t}$ (2.16), is a Markov process on E^N with homogeneous infinitesimal generator $\overline{L}^N$ defined by

$$\overline{L}^N(F)(x^1, \ldots, x^N) := \sum_{i=1}^{N} \overline{\mathcal{L}}^{(i)}_{m(\underline{x})}(F)(x^1, \ldots, x^i, \ldots, x^N) \ , \tag{4.1}$$

for any $F \in \mathcal{C}_b(E^N)$. Here $\overline{\mathcal{L}}^{(i)}_{m(\underline{x})}$ denotes the McKean generator $\overline{\mathcal{L}}_{m(\underline{x})}$ (2.16) acting

on the function $x^i \mapsto F(x^1, \dots, x^i, \dots, x^N)$, where the dependence on μ has been replaced by the empirical distribution $m(\underline{x})$.

In analogy to the decomposition $\overline{\mathcal{L}}_\mu = \mathcal{L} + \widetilde{L}_\mu$ in (2.16), the generator (4.1) can be decomposed as $\overline{L}^N = L^N + \widetilde{L}^N$ with

$$L^N(F)(\underline{x}) := \sum_{i=1}^N \mathcal{L}^{(i)}(F)(\underline{x}) \ , \tag{4.2}$$

$$\widetilde{L}^N(F)(\underline{x}) := \sum_{i=1}^N \widetilde{\mathcal{L}}^{(i)}_{m(\underline{x})}(F)(\underline{x}) \ , \tag{4.3}$$

where $\mathcal{L}^{(i)}$ and $\widetilde{\mathcal{L}}^{(i)}_{m(\underline{x})}$ stand respectively for the operators $\mathcal{L}$ and $\widetilde{\mathcal{L}}_{m(\underline{x})}$ acting on the function $x^i \mapsto F(\underline{x})$, i.e. only on particle i.

Moreover, using representation (2.18) for $\widetilde{\mathcal{L}}_\mu$, we can write

$$\widetilde{\mathcal{L}}^{(i)}_{m(\underline{x})}(F)(\underline{x}) \ = \ \frac{1}{N} \sum_{j=1}^N \widetilde{W}(x_i, \, x_j) \left(F(\underline{x}^{i,x_j}) - F(\underline{x}) \right) \ , \tag{4.4}$$

with $\underline{x}^{i,y} := (x_1, \dots, x_{i-1}, y, x_{i+1}, \dots, x_N)$, which introduces an interaction between the particles. In this decomposition, (4.2) generates the so-called *mutation dynamics*, where the particles evolve independently under the dynamics given by the infinitesimal generator $\mathcal{L}$ of the original process, whereas (4.3) generates the *selection dynamics*, which leads to mean field interactions between particles. With (4.4), for every particle $i = 1, \dots, N$, the state of i gets killed and replaced by that of particle j with overall escape rate $\frac{1}{N}\widetilde{W}(x_i, \, x_j)$. Hence, the total selection rate for a particle i is $\frac{1}{N}\sum_{j=1}^N \widetilde{W}(x_i, \, x_j)$, and depends on the McKean representation, in particular the choice of $\widetilde{\mathcal{L}}_\mu$ in (2.18).

4.1.1 Convergence results

By definition of $\overline{L}^N$ (4.1), for any function F on E^N of the form $F(\underline{x}) = m(\underline{x})(f)$, with $f \in \mathcal{C}_b(E)$, we have that

$$\overline{L}^N(F)(\underline{x}) \ = \ m(\underline{x})\big(\overline{\mathcal{L}}_{m(\underline{x})}(f)\big) \ , \tag{4.5}$$

$$\Gamma_{\overline{L}^N}(F, \, F)(\underline{x}) \ = \ \frac{1}{N} m(\underline{x})\big(\Gamma_{\overline{\mathcal{L}}_{m(\underline{x})}}(f, \, f)\big) \ , \tag{4.6}$$

thus conditions (3.3a)-(3.3b) are satisfied.

Analogous relations hold also for the individual mutation and cloning parts of the generator. Since generators are linear, the identity (4.5) is immediate. The carré

du champ (4.6) is quadratic in F, but off-diagonal terms in the corresponding double sum turn out to vanish in a straightforward computation, leading to the additional factor $1/N$. More in detail, using the Definition 2.1.6 of carré du champ,

$$\Gamma_{\overline{L}^N}(F,\,F)(\underline{x}) \;=\; \sum_{i=1}^{N} \overline{\mathcal{L}}^{(i)}_{m(\underline{x})}(F^2)(\underline{x}) \;-\; 2\,F(\underline{x}) \sum_{i=1}^{N} \overline{\mathcal{L}}^{(i)}_{m(\underline{x})}(F)(\underline{x}) \;,$$

and observing that, for any F of the form $F(\underline{x}) = m(\underline{x})(f)$, we have

$$\sum_{i=1}^{N} \overline{\mathcal{L}}^{(i)}_{m(\underline{x})}(F^2)(\underline{x}) \;=\; \frac{1}{N^2}\Big(\sum_{i=1}^{N} \overline{\mathcal{L}}_{m(\underline{x})}(f^2)(x_i) + 2\sum_{i=1}^{N} \overline{\mathcal{L}}_{m(\underline{x})}(f)(x_i)\cdot\sum_{j\neq i} f(x_j)\Big) \;,$$

and

$$F(\underline{x})\sum_{i=1}^{N} \overline{\mathcal{L}}^{(i)}_{m(\underline{x})}(F)(\underline{x}) = \frac{1}{N^2}\Big(\sum_{i=1}^{N} f(x_i)\,\overline{\mathcal{L}}_{m(\underline{x})}(f)(x_i) + \sum_{i=1}^{N} \overline{\mathcal{L}}_{m(\underline{x})}(f)(x_i)\cdot\sum_{j\neq i} f(x_j)\Big) \;,$$

we obtain the identity (4.6).

Remark. Note that similar relations hold also for marginal test functions $F(\underline{x}) = f(x_l)$, $f \in \mathcal{C}_b(E)$, depending only on a single particle, namely

$$\overline{L}^N(F)(\underline{x}) \;=\; \overline{\mathcal{L}}_{m(\underline{x}}(f)(x_l) \quad \text{and} \quad \Gamma_{\overline{L}^N}(F,F)(\underline{x}) \;=\; \Gamma_{\overline{\mathcal{L}}_{m(\underline{x}}}(f,f)(x_l) \;.$$

So, generator and carré du champ both coincide with the corresponding operators for the McKean dynamics. This means that for large enough N and $\mu^N(\xi_t)$ close to μ_t, each marginal process $t \mapsto \xi^l_t$ has essentially the same distribution as the corresponding McKean process $t \mapsto \xi^l_t$. However, while the marginal processes for a given particle approximation are identically distributed, they are not independent due to the correlations between particles resulting from selection events.

Furthermore, by construction, for almost every realisation ξ_t, $t > 0$, of the mean field particle approximation, there exists at most one particle i such that $\xi^i_t \neq \xi^i_{t-}$, since during mutation events, as well as selection events, only one particle changes its state. In other words,

$$\sup_{t\geq 0} \big|\{i \in 1,\ldots,N : \xi^i_t \neq \xi^i_{t-}\}\big| \;\leq\; 1 \;,$$

and thus condition (3.3c) is satisfied with $K = 1$. Therefore, under the initial condition (3.3d), Theorem 3.1.2 holds and provides L^p-error and bias estimates of order $1/\sqrt{N}$ and $1/N$ respectively, in accordance with already established results,

e.g. in Rousset (2006); Del Moral and Miclo (2003); Del Moral (2013).

Proposition 4.1.1 (CLT). *Let μ_t^N be the empirical measure of the mean field N-particle model associated to a McKean interpretation $\overline{\mathcal{L}}_{\mu_t}$ (2.15) of the Feynman-Kac marginal μ_t such that the operator $\widetilde{W}(x,y)$ defining the selection generator $\widetilde{\mathcal{L}}_\mu$ (2.18) satisfies*

$$\lim_{z\to x} \|\widetilde{W}(z,\cdot) - \widetilde{W}(x,\cdot)\| = 0 \,. \tag{4.7}$$

Then, under the initial condition (3.3d), for any fixed $T>0$, the sequence

$$V_T^N(f) := \sqrt{N}\big(\mu_T^N(f) - \mu_T(f)\big)$$

converges in law as $N\to\infty$ to a centered Gaussian random variable $V_T(f)$ with variance given by

$$\begin{aligned}\mathbb{E}\big[V_T(f)^2\big] =& \mu_T\big((f-\mu_T(f))^2\big) \\ &+ \int_0^T \mu_s\Big((\Theta_{s,T}\overline{f})^2\ \big(\mathcal{V} - \mu_s(\mathcal{V})\big) + \Gamma_{\widetilde{\mathcal{L}}_{\mu_s}}(\Theta_{s,T}\overline{f},\,\Theta_{s,T}\overline{f})\Big)\,ds\,.\end{aligned} \tag{4.8}$$

Proof. For any $f\in\mathcal{C}_b(E)$, $G_\mu(f,f)=\Gamma_{\overline{\mathcal{L}}_\mu}(f,f)$ is in the form (3.32) with

$$\begin{aligned} g_0(x) &= \Gamma_{\mathcal{L}}(f,f)(x)\,, \\ g_1(y,x) &= \widetilde{W}(x,y)\,\big(f(y)-f(x)\big)^2\,, \\ g_2(y,x) &= 0\,. \end{aligned}$$

Therefore, condition (3.33) is satisfied and we can apply Lemma 3.3.2 and obtain that the Central Limit Theorem (Theorem 3.1.3) holds. In particular,

$$\mathbb{E}\big[V_T(f)^2\big] = \mu_0\big((\Theta_{0,T}\overline{f})^2\big) + \int_0^T \mu_s\Big(\Gamma_{\overline{\mathcal{L}}_{\mu_s}}(\Theta_{s,T}\overline{f},\,\Theta_{s,T}\overline{f})\Big)\,ds. \tag{4.9}$$

To conclude, it is enough to observe that, applying the evolution equation (2.14) of μ_t and then Lemma 3.2.2, we have

$$\begin{aligned}&\mu_t((\Theta_{t,T}\overline{f})^2) - \mu_0((\Theta_{0,T}\overline{f})^2) \\ &= \int_0^t \mu_s\big(2\,\Theta_{s,T}\overline{f}\cdot\partial_s(\Theta_{s,T}\overline{f}) + \mathcal{L}(\Theta_{s,T}\overline{f})^2 + (\Theta_{s,T}\overline{f})^2\cdot(\mathcal{V}-\mu_s(\mathcal{V}))\big)\,ds \\ &= \int_0^t \mu_s\big(-2\Theta_{s,T}\overline{f}\cdot\mathcal{L}(\Theta_{s,T}\overline{f}) + \mathcal{L}(\Theta_{s,T}\overline{f})^2 + (\Theta_{s,T}\overline{f})^2\cdot\big(\mu_s(\mathcal{V})-\mathcal{V}\big)\big)\,ds \\ &= \int_0^t \mu_s\big(\Gamma_{\mathcal{L}}(\Theta_{s,T}\overline{f},\,\Theta_{s,T}\overline{f}) + (\Theta_{s,T}\overline{f})^2\cdot\big(\mu_s(\mathcal{V})-\mathcal{V}\big)\big)\,ds.\end{aligned}$$

Thus,

$$\int_0^t \mu_s\Big(\Gamma_{\mathcal{L}}(\Theta_{s,T}\overline{f}, \Theta_{s,T}\overline{f})\Big)\, ds$$
$$= \mu_t((\Theta_{t,T}\overline{f})^2) - \mu_0((\Theta_{0,T}\overline{f})^2) + \int_0^t \mu_s\Big((\Theta_{s,T}\overline{f})^2 \cdot \big(\mathcal{V} - \mu_s(\mathcal{V})\big)\Big)\, ds\ . \tag{4.10}$$

Substituting into the expression (4.9), we obtain the statement.

□

Remark. Condition (4.7) is just a technical regularity condition on $\widetilde{W}$ which imposes a degree of continuity on G, namely

$$\lim_{(z,N)\to(x,\infty)} G_{\mu_t^N}(f,f)(z) \;=\; G_{\mu_t}(f,f)(x) \qquad \mathbb{P}_{\mu_0} - \text{a.s.}$$

for any $x \in E$, so to ensure the a.s. convergence for the sequence $\mu_t^N\big(G_{\mu_t^N}(f,f)\big)$. Moreover, observe that condition (4.7) is satisfied for all the three McKean models considered in Section 2.3. Indeed, it is immediate to see that the limit (4.7) is satisfied for (2.21a) and (2.21c) since $\mathcal{V} \in \mathcal{C}_b(E)$. Now, consider the McKean model defined by (2.21b). For any $x \in E$ and $\varepsilon > 0$, there exists $\delta > 0$ such that, for any $z \in E$ s.t. $|z - x| < \delta$, $|\mathcal{V}(x) - \mathcal{V}(z)| < \varepsilon$. Hence, for any $y \in E$ and $z \in E$ s.t. $|z - x| < \delta$, we have the following bounds:

$$\big(\mathcal{V}(y) - \mathcal{V}(z)\big)^+ \;\le\; \big(\mathcal{V}(y) - \mathcal{V}(x) + \varepsilon\big)^+ \;\le\; \big(\mathcal{V}(y) - \mathcal{V}(x)\big)^+ + \varepsilon\ ,$$
$$\big(\mathcal{V}(y) - \mathcal{V}(z)\big)^+ \;\ge\; \big(\mathcal{V}(y) - \mathcal{V}(x) - \varepsilon\big)^+ \;\ge\; \big(\mathcal{V}(y) - \mathcal{V}(x)\big)^+ - \varepsilon\ .$$

In conclusion, for any $z \in E$ s.t. $|z - x| < \delta$,

$$\big|\widetilde{W}(x,y) \;-\; \widetilde{W}(z,y)\big| \;\le\; \varepsilon,$$

independently of the choice of $y \in E$ and, thus, condition (4.7) is satisfied.

4.1.2 Some remarks on the asymptotic variance

From general practical experience it is favourable to minimise the total selection rate in order to improve the estimator's asymptotic variance; indeed it's widely understood in the SMC literature that eliminating unnecessary selection events can significantly improve estimator variances (see, for example, Gerber et al., 2019). This can be explained by the fact that the selection events in a particle approximation

increase the correlations among the particles in the ensemble, and thereby decrease the resolution in the empirical distribution μ_t^N. For mean field particle approximations this suggests that (2.21b) is preferable to (2.21a) since the total selection rate for the McKean model (2.21a) is

$$S_c^1(\underline{x}) := \frac{1}{N}\sum_{i,j=1}^{N}\Big(\big(\mathcal{V}(x_i)-c\big)^- + \big(\mathcal{V}(x_j)-c\big)^+\Big) = \sum_{i=1}^{N}\big|\mathcal{V}(x_i)-c\big|\ ,$$

by symmetry of summations, whereas the total selection rate for the McKean model (2.21b) is

$$S^2(\underline{x}) := \frac{1}{N}\sum_{i,j=1}^{N}\big(\mathcal{V}(x_i)-\mathcal{V}(x_j)\big)^+ = \frac{1}{2N}\sum_{i,j=1}^{N}\big|\mathcal{V}(x_i)-\mathcal{V}(x_j)\big|\ ,$$

hence, by triangular inequality, we can see that $S_c^1(\underline{x}) \geq S^2(\underline{x})$, for every $c \in \mathbb{R}$. The fact that the McKean model (2.21b) leads to a lower total selection rate compared to the McKean model (2.21a) can also be seen more intuitively by observing that, using (2.21b), every selection event increases the fitness potential $\mathcal{V}$, whereas this is not necessarily the case with (2.21a), which increases the potential only on average. From the point of view of the total selection rate, there are even more optimal choices than (2.21b), such as the McKean model (2.21c), for instance. Indeed, the total selection rate for the McKean model (2.21c) is

$$\begin{aligned} S^3(\underline{x}) &= \frac{1}{N}\sum_{i,j=1}^{N}\big(\mathcal{V}(x_i)-m(\underline{x})(\mathcal{V})\big)^-\frac{\big(\mathcal{V}(x_j)-m(\underline{x})(\mathcal{V})\big)^+}{m(\underline{x})\big((\mathcal{V}-m(\underline{x})(\mathcal{V}))^+\big)} \\ &= \sum_{i=1}^{N}\big(\mathcal{V}(x_i)-m(\underline{x})(\mathcal{V})\big)^- = \frac{1}{2}\sum_{i=1}^{N}\big|\mathcal{V}(x_i)-m(\underline{x})(\mathcal{V})\big|\ , \end{aligned}$$

hence $S^3(\underline{x}) \leq S^2(\underline{x})$, by applying the Jensen's inequality to the convex function $v \mapsto |\mathcal{V}(x_i)-v|$, for any given $i = 1, \ldots, N$.

On the other hand, depending on the particular application, implementing particle approximations with lower total selection rate could be computationally more expensive, leading to a trade-off in lower values for N to be accessible in practice, as illustrated in Section 5.5, for the particular example of inclusion processes.

More precisely, to improve the accuracy of an algorithm, we need to reduce the asymptotic variance. In the context of mean field particle approximations, if the hypotheses of Proposition 4.1.1 are satisfied, the asymptotic variance $\mathbb{E}\big[V_T(f)^2\big]$ associated to a mean field particle approximation can be written in the form (4.8).

Unfortunately, we see that $\mathbb{E}\big[V_T(f)^2\big]$ cannot be estimated a priori, as it depends on the (normalised) Feynman-Kac measures μ_t and on the corresponding propagator $\Theta_{t,T}$, with $t \in [0,T]$, thus we cannot provide a quantitative comparison between mean field particle approximations based on different McKean models. However, we can still obtain some preliminary qualitative results.

Lemma 4.1.2. *Consider the mean field N-particle systems associated to two different McKean models $\overline{\mathcal{L}}^a_{\mu_T}$ and $\overline{\mathcal{L}}^b_{\mu_T}$ in the form (2.16). Then, using the same notation of Proposition 4.1.1 and specifying the upper-indices a and b to distinguish the quantities associated respectively to $\overline{\mathcal{L}}^a_{\mu_T}$ and $\overline{\mathcal{L}}^b_{\mu_T}$,*

$$\begin{aligned}\mathbb{E}\big[V^a_T(f)^2\big] \,-\, \mathbb{E}\big[V^b_T(f)^2\big] \\ = \int_0^T \mu_s\big(\, \Gamma_{\widetilde{\mathcal{L}}^a_{\mu_s}}(\Theta_{s,T}\overline{f},\, \Theta_{s,T}\overline{f}) \,-\, \Gamma_{\widetilde{\mathcal{L}}^b_{\mu_s}}(\Theta_{s,T}\overline{f},\, \Theta_{s,T}\overline{f})\,\big)\, ds\,.\end{aligned}$$

Proof. It is a direct consequence of Central Limit Theorem for mean field particle approximations (Proposition 4.1.1).

□

Lemma 4.1.3. *Consider the class of McKean generators $\widetilde{\mathcal{L}}_{\mu,c}$ defined by (2.21a), with kernel*

$$\widetilde{W}_c(x,y)\,\mu(dy) \,=\, \Big(\big(\mathcal{V}(x)-c\big)^- \,+\, \big(\mathcal{V}(y)-c\big)^+\Big)\,\mu(dy)\;,$$

with $c \in \mathbb{R}$ possibly depending on the current distribution μ of the process. Given $f \in \mathcal{C}_b(E)$ and $\mu \in \mathcal{P}(E)$, the value of c that minimises $\mu\big(\Gamma_{\widetilde{\mathcal{L}}_{\mu,c}}(f,f)\big)$ depends on f and μ, is the real number $c \in \mathbb{R}$ such that

$$\int_{x\in E}\int_{\mathcal{V}(y)\geq c}\big(f(y)-f(x)\big)^2\mu(dy)\mu(dx) \,\geq\, \frac{I}{2}\;, \tag{4.11}$$

and

$$\int_{x\in E}\int_{\mathcal{V}(y)\leq c}\big(f(y)-f(x)\big)^2\mu(dy)\mu(dx) \,\geq\, \frac{I}{2}\;, \tag{4.12}$$

where $I = \int_{E^2}\big(f(y)-f(x)\big)^2\mu(dy)\mu(dx)$.

Proof. It is enough to observe that

$$\begin{aligned}\mu\big(\Gamma_{\widetilde{\mathcal{L}}_{\mu,c}}(f,f)\big) &= \int_{\mathcal{V}(x)\le c}\int_{y\in E}\big(c-\mathcal{V}(x)\big)\big(f(y)-f(x)\big)^2\,\mu(dy)\mu(dx)\\ &\quad+\int_{x\in E}\int_{\mathcal{V}(y)\ge c}\big(\mathcal{V}(y)-c\big)\big(f(y)-f(x)\big)^2\,\mu(dy)\mu(dx)\\ &=\int_{x\in E}\int_{y\in E}\big|\mathcal{V}(y)-c\big|\big(f(x)-f(y)\big)^2\,\mu(dy)\mu(dx)\ ,\end{aligned}$$

which is well-known that it is minimised by the real number $c\in\mathbb{R}$ such that (4.11) and (4.12) hold (see e.g. Cramér, 1999, p. 179).

□

Lemma 4.1.4. *Let $\mathcal{K}$ be the class of infinitesimal 'selection' generators $\widetilde{\mathcal{L}}_\mu$, $\mu\in\mathcal{P}(E)$, in the form (2.18) with $\widetilde{W}(x,y)\ge 0$ and such that condition (2.20) holds, that is*

$$\widetilde{W}(y,x)\;-\;\widetilde{W}(x,y)\;=\;\mathcal{V}(x)\;-\;\mathcal{V}(y)\ ,$$

for any $x,y\in E$. Denote by $\widetilde{\mathcal{L}}^{\mathcal{V}}_\mu$ the infinitesimal generator for the selection dynamics associated to the McKean model defined by (2.21b), with kernel

$$\widetilde{W}(x,y)\,\mu(dy)\;=\;\big(\mathcal{V}(y)-\mathcal{V}(x)\big)^+\,\mu(dy)\ .$$

Then, $\widetilde{\mathcal{L}}^{\mathcal{V}}_\mu$ is the optimal generator in $\mathcal{K}$, i.e. $\widetilde{\mathcal{L}}^{\mathcal{V}}_\mu\in\mathcal{K}$ and

$$\mu\big(\Gamma_{\widetilde{\mathcal{L}}^{\mathcal{V}}_\mu}(f,f)\big)\;=\;\min_{\widetilde{\mathcal{L}}_\mu\in\mathcal{K}}\mu\big(\Gamma_{\widetilde{\mathcal{L}}_\mu}(f,f)\big)\ ,$$

for any $\mu\in\mathcal{P}(E)$ and $f\in\mathcal{C}_b(E)$.

Remark. Note that the infinitesimal generators given by (2.21a) and (2.21b) satisfy condition (2.20), hence they are in $\mathcal{K}$, whereas the McKean model (2.21c) does not satisfy the pointwise assumption (2.20).

Proof. First, it is easy to see that $\widetilde{\mathcal{L}}^{\mathcal{V}}_\mu$ satisfies condition (2.20), hence $\widetilde{\mathcal{L}}^{\mathcal{V}}_\mu\in\mathcal{K}$.

Given $\mu\in\mathcal{P}(E)$ and $f\in\mathcal{C}_b(E)$, note that for any 'selection' generator $\widetilde{\mathcal{L}}_\mu$ we can write

$$\mu\big(\Gamma_{\widetilde{\mathcal{L}}_\mu}(f,f)\big)\;=\;\frac{1}{2}\int_{E^2}\big(\widetilde{W}(x,y)+\widetilde{W}(y,x)\big)\big(f(y)-f(x)\big)^2\,\mu(dy)\mu(dx)\ .$$

Given $x,y\in E$, the minimal value of $\big(\widetilde{W}(x,y)+\widetilde{W}(y,x)\big)$ subject to (2.20) and

$\widetilde{W}(x,y)$, $\widetilde{W}(y,x) \geq 0$ is achieved by

$$\begin{aligned} \widetilde{W}^\star(y,x) &= \mathcal{V}(x) - \mathcal{V}(y) \quad \text{and} \quad \widetilde{W}^\star(x,y) = 0 \ , \quad \text{if } \mathcal{V}(x) \geq \mathcal{V}(y) \ , \\ \widetilde{W}^\star(y,x) &= 0 \quad \text{and} \quad \widetilde{W}^\star(x,y) = \mathcal{V}(y) - \mathcal{V}(x) \ , \quad \text{if } \mathcal{V}(y) \geq \mathcal{V}(x) \ . \end{aligned}$$

We conclude by noting that $\widetilde{W}^\star(x,y) = \big(\mathcal{V}(y) - \mathcal{V}(x)\big)^+$ for any $x, y \in E$.

□

Remark. Lemma 4.1.4 implies in particular that, for any $f \in \mathcal{C}_b(E)$, $\mu \in \mathcal{P}(E)$ and for every choice of $c \in \mathbb{R}$,

$$\mu\big(\Gamma_{\widetilde{\mathcal{L}}^{\mathcal{V}}_{\mu}}(f,f)\big) \leq \mu\big(\Gamma_{\widetilde{\mathcal{L}}_{\mu,c}}(f,f)\big) \ ,$$

where $\widetilde{\mathcal{L}}_{\mu,c}$ is the selection generator considered in Lemma 4.1.3, and the inequality is strict except for degenerate cases, e.g. if $\mathcal{V}$ takes only two values and c lies in between the two.

4.2 A generalised cloning algorithm

Cloning algorithms have been proposed in the theoretical physics literature (Giardinà et al., 2006; Lecomte and Tailleur, 2007) for evaluating large deviation functions associated to Markov processes similar to the mean field system (4.1), using the same mutation dynamics. While selection and mutation events are independent in the latter due to the additive structure of $\overline{L}^N$ in (4.2) and (4.3), in cloning algorithms both are combined to reduce computational cost. Recalling that the overall escape rate and probability kernel of the original dynamics $\mathcal{L}$ are denoted respectively by $\lambda(x)$ and $p(x, dy)$, the dynamics of the cloning algorithm can be described as follows:

- Any particle $i = 1, \ldots, N$ "jumps" (i.e. it changes its state) with rate $\lambda(x_i)$ depending only on the state of the given particle but not on the rest of the population;
- During a jump of a particle i, a set A of particles is chosen at random from the ensemble with probability $\pi_{\underline{x}}(x_i, A)$ possibly dependent on the whole population $\underline{x} \in E^N$, and every particle $j \in A$ is replaced by a clone of i;
- Then, the particle i mutates to a new state $y \in E$ with probability kernel $p(x_i, dy)$, so that the overall escape rate at which a clone i mutates to a new state y is $W(x_i, dy)$.

Hence, the infinitesimal description of the cloning process is given by the generator

$$\overline{L}^N(F)(\underline{x}) = \sum_{i=1}^{N} \lambda(x_i) \sum_{A\in\mathcal{N}} \int_{y\in E} \pi_{\underline{x}}(x_i, A) \cdot p(x_i, dy) \left(F(\underline{x}^{A,x_i;\,i,y}) - F(\underline{x})\right), \quad (4.13)$$

for any $F \in \mathcal{C}_b(E^N)$ and $\underline{x} \in E^N$, where $\mathcal{N}$ is the set of all subsets of N particle indices and $\underline{x}^{A,w;\,i,y}$ denotes the vector $(z_1, \dots, z_N) \in E^N$ with

$$z_j := \begin{cases} x_j & j \notin A,\ j \neq i \\ w & j \in A,\ j \neq i \\ y & j = i, \end{cases}$$

for $j \in \{1, \dots, N\}$ and $w, y \in E$.

4.2.1 Convergence results

Let $\psi_{\underline{x}}(x_i, x_j)$ be the probability of replacing a particle $j \in \{1, \dots, N\}$ during a cloning event for a particle $i = 1, \dots, N$, i.e.

$$\psi_{\underline{x}}(x_i, x_j) := \sum_{A|j\in A} \pi_{\underline{x}}(x_i, A) \ ,$$

where the summation is over all sets $A \in \mathcal{N}$ containing the index j.

The following result shows that condition (3.3a) on the generator, which ensures the correct expected behaviour of mean-field observables, is satisfied for the cloning algorithm if $\psi_{\underline{x}}$ is in the form

$$\psi_{\underline{x}}(x_i, x_j) = \frac{\widetilde{W}(x_j, x_i)}{N \cdot \lambda(x_i)} \ , \quad N \geq \sup_{x,y\in E} \frac{\widetilde{W}(y, x)}{\lambda(x)} \ , \quad (4.14)$$

for every i, $j \in \{1, \dots, N\}$, where $\widetilde{W}(x, y)$ is the kernel defining the McKean model (2.16)-(2.18).

Proposition 4.2.1. *Given a McKean interpretation $\overline{\mathcal{L}}_\mu$ (2.16)-(2.18), let $\overline{L}^N$ be the generator (4.13) associated to a cloning process satisfying (4.14), then*

$$\overline{L}^N(F)(\underline{x}) = m(\underline{x})\big(\overline{\mathcal{L}}_{m(\underline{x})}(f)\big) \ ,$$

for any test function of the form $F(\underline{x}) = m(\underline{x})(f)$, $f \in \mathcal{C}_b(E)$, and $\underline{x} \in E^N$.

Proof. We start by considering the first term in the expression of $\overline{L}^N$ (4.13). Observe that with $F(\underline{x}) = m(\underline{x})(f)$,

$$\begin{aligned} F(\underline{x}^{A,x_i;\,i,y}) - F(\underline{x}) \;&=\; \frac{1}{N}\big(f(y) - f(x_i)\big) \;+\; \frac{1}{N}\sum_{j\in A}\big(f(x_i) - f(x_j)\big) \\ &=\; \big(F(\underline{x}^{i,y}) - F(\underline{x})\big) \;+\; \big(F(\underline{x}^{A,x_i}) - F(\underline{x})\big)\;. \end{aligned} \tag{4.15}$$

Thus, we can write

$$\begin{aligned} &\int_{y\in E} p(x_i,\,dy)\sum_{A\in\mathcal{N}} \pi_{\underline{x}}(x_i,\,A)\,\big(F(\underline{x}^{A,x_i;\,i,y}) - F(\underline{x})\big) \\ &\quad= \int_{y\in E} p(x_i,\,dy)\sum_{A\in\mathcal{N}} \pi_{\underline{x}}(x_i,\,A)\,\Big(\big(F(\underline{x}^{i,y}) - F(\underline{x})\big) + \big(F(\underline{x}^{A,x_i}) - F(\underline{x})\big)\Big) \\ &\quad= \int_{E} p(x_i,dy)\big(F(\underline{x}^{i,y}) - F(\underline{x})\big) \;+\; \sum_{A\in\mathcal{N}} \pi_{\underline{x}}(x_i,\,A)\,\big(F(\underline{x}^{A,x_i}) - F(\underline{x})\big)\;. \end{aligned}$$

Moreover,

$$\begin{aligned} \sum_{A\in\mathcal{N}} \pi_{\underline{x}}(x_i,\,A)\,\big(F(\underline{x}^{A,x_i}) - F(\underline{x})\big) \;&=\; \sum_{j=1}^{N}\sum_{A\,|\,j\in A} \pi_{\underline{x}}(x_i, A)\,\big(f(x_i) - f(x_j)\big) \\ &=\; \frac{1}{N}\sum_{j=1}^{N} \frac{\widetilde{W}(x_j, x_i)}{\lambda(x_i)}\,\big(f(x_i) - f(x_j)\big)\;, \end{aligned} \tag{4.16}$$

by hypothesis. As a result, the generator $\overline{L}^N$ (4.13) can be rewritten as

$$\begin{aligned} \overline{L}^N(F)(\underline{x}) \;=& \frac{1}{N}\sum_{i=1}^{N}\lambda(x_i)\int_E p(x_i,\,dy)\big(f(y) - f(x_i)\big) \\ &+ \frac{1}{N^2}\sum_{i,j=1}^{N}\widetilde{W}(x_j, x_i)\,\big(f(x_j) - f(x_i)\big) \\ &= m(\underline{x})\big(\overline{\mathcal{L}}_{m(\underline{x})}(f)\big) \end{aligned}$$

by changing summation variables in the cloning term.

□

Remark. Note that a similar relation holds also for marginal test functions $F(\underline{x}) = f(x_l)$, $f \in \mathcal{C}_b(E)$, $l \in \{1, \ldots, N\}$, depending only on a single particle, namely

$$\overline{L}^N(F)(\underline{x}) \;=\; \overline{\mathcal{L}}_{m(\underline{x})}(f)(x_l)\;.$$

Indeed,

$$\begin{aligned}\overline{L}^N(F)(\underline{x}) \;&=\; \lambda(x_l)\int_{y\in E} p(x_l,\,dy)\,\big(f(y)-f(x_l)\big)\\ &\qquad+\sum_{i=1}^{N}\lambda(x_i)\,\psi_{\underline{x}}(x_i,\,x_l)\,\big(f(x_i)-f(x_l)\big)\\ &=\;\overline{\mathcal{L}}_{m(\underline{x})}f(x_l)\;,\end{aligned}$$

where the last equality follows by condition (4.14). Moreover, since this holds for every marginal test function, we also obtain

$$\begin{aligned}\Gamma_{\overline{L}^N}(F,F)(\underline{x}) \;&=\; \overline{L}^N(F^2)(\underline{x})\;-\;2\,F(\underline{x})\cdot\overline{L}^N(F)(\underline{x})\\ &=\;\Gamma_{\overline{\mathcal{L}}_{m(\underline{x})}}(f,f)(x_l)\;.\end{aligned}$$

So, as for mean field particle systems, the generator and carré du champ of a cloning process satisfying (4.14) both coincide with the corresponding operators for the McKean dynamics. This means that for large enough N and $\mu^N(\xi_t)$ close to μ_t, each marginal process $t\mapsto\xi_t^l$ has essentially the same distribution as the corresponding McKean process $t\mapsto\xi_t^l$. However, while the marginal processes for a given particle approximation are identically distributed, obviously they are not independent due to the correlations between particles resulting from cloning events.

To control the fluctuations of the cloning algorithm we introduce the probability $\Psi_{\underline{x}}(x_i;\,x_j,x_k)$ of replacing two different particles $j,k\in\{1,\dots,N\}$, $j\neq k$, during a cloning event for a particle $i=1,\dots,N$, namely

$$\Psi_{\underline{x}}(x_i;\,x_j,x_k):=\sum_{A|j,k\in A}\pi_{\underline{x}}(x_i,A)\;,$$

for any i, j, $k\in\{1,\dots,N\}$. The following result shows that condition (3.3b) on the carré du champ, which ensures that the trajectories concentrate on their expectation, is satisfied for the cloning algorithm in case $\Psi_{\underline{x}}$ can be written in the form

$$\Psi_{\underline{x}}(x_i;\,x_j,\,x_k)\;=\;\frac{Q(x_i)}{N^2}\;\omega(x_i,x_j)\cdot\omega(x_i,x_k)+O\Big(\frac{1}{N^3}\Big)\;,\tag{4.17}$$

for any $i,j,k\in\{1,\dots,N\}$, $j\neq k$, where the functions $Q\in\mathcal{C}_b(E)$ and $\omega\in\mathcal{C}_b(E^2)$ are independent of N.

Proposition 4.2.2. *Given a McKean interpretation $\overline{\mathcal{L}}_\mu$ (2.16)-(2.18), let $\overline{L}^N$ be the*

generator (4.13) *associated to a cloning process satisfying the hypothesis in Proposition 4.2.1 and such that* $\Psi_{\underline{x}}$ *satisfies condition* (4.17).

Then, for any test function of the form $F(\underline{x}) = m(\underline{x})(f)$, *with* $f \in \mathcal{C}_b(E)$,

$$\Gamma_{\overline{L}^N}(F,\, F)(\underline{x}) \;=\; \frac{1}{N}\, m(\underline{x})\Big(G_{m(\cdot)}(f,f)\Big) \;+\; O\Big(\frac{1}{N^2}\Big)\;,$$

as $N \to \infty$, *where*

$$\begin{aligned} G_\mu(f,f)(x) \;=& \Gamma_{\overline{\mathcal{L}}_\mu}(f,f)(x) \;+\; \lambda(x)\cdot Q(x)\,\big(\ell_\mu(f)(x)\big)^2 \\ &- \frac{2}{\lambda(x)}\mathcal{L}(f)(x)\cdot\widetilde{\mathcal{L}}^t_\mu(f)(x)\;, \end{aligned} \tag{4.18}$$

with

$$\ell_\mu(f)(x) := \int_E \omega(x,\, y)\,\big(f(y) - f(x)\big)\,\mu(dy)\;,$$

and

$$\widetilde{\mathcal{L}}^t_\mu(f)(x) := \int_E \widetilde{W}(y,x)\,\big(f(y) - f(x)\big)\,\mu(dy)\;.$$

Remark. Due to the linearity of the generator, the combined mutation/cloning events in the cloning algorithm can be decomposed easily, which leads to extra terms only in the quadratic carré du champ. In the expression of the operator G_μ (4.18), the term

$$\frac{1}{\lambda(x)}\mathcal{L}(f)(x)\cdot\widetilde{\mathcal{L}}^t_\mu(f)(x)$$

is due to the dependence between mutation and cloning dynamics and its sign is not known a priori. Whereas, the term $\lambda(x)\cdot Q(x)\,\big(\ell_\mu(f)(x)\big)^2$ arises from the dependence between clones (since multiple cloning events are allowed at the same time) and is always non-negative. In particular, in any setting in which there is at most one clone per event, i.e. when $\Psi_{\underline{x}} \equiv 0$, the term vanishes.

Moreover, note that the generator $\widetilde{\mathcal{L}}^t_\mu(f)$ differs from the selection generator $\widetilde{\mathcal{L}}_\mu(f)$, having kernel $\widetilde{W}(y,x)\mu(dy)$ instead of $\widetilde{W}(x,y)\mu(dy)$.

Proof. Consider the carré du champ of $\overline{L}^N$,

$$\begin{aligned} &\Gamma_{\overline{L}^N}\big(F,\, F\big)(\underline{x}) \\ &\quad= \sum_{i=1}^N \lambda(x_i)\int_E p(x_i,\, dy)\sum_{A\in\mathcal{N}}\pi_{\underline{x}}(x_i,\, A)\cdot\Big(F(\underline{x}^{A,x_i;\,i,y}) - F(\underline{x})\Big)^2\;. \end{aligned}$$

Using the decomposition (4.15), we can write

$$\begin{aligned}
&\int_E p(x_i, dy) \sum_{A \in \mathcal{N}} \pi_{\underline{x}}(x_i,\, A)\big(F(\underline{x}^{A,x_i;i,y}) - F(\underline{x})\big)^2 \\
&\quad = \int_E p(x_i, dy)\big(F(\underline{x}^{i,y}) - F(\underline{x})\big)^2 + \sum_{A \in \mathcal{N}} \pi_{\underline{x}}(x_i,\, A)\big(F(\underline{x}^{A,x_i}) - F(\underline{x})\big)^2 \\
&\quad + 2\int_E p(x_i, dy)\ \big(F(\underline{x}^{i,y}) - F(\underline{x})\big) \sum_{A \in \mathcal{N}} \pi_{\underline{x}}(x_i,\, A)\big(F(\underline{x}^{A,x_i}) - F(\underline{x})\big)\ ,
\end{aligned}$$

where, by (4.16), the last line can be rewritten as

$$\begin{aligned}
&\frac{2}{N^2}\int_E p(x_i, dy)\ \big(f(y) - f(x_i)\big) \sum_{j=1}^N \frac{\widetilde{W}(x_j, x_i)}{\lambda(x_i)}\ \big(f(x_i) - f(x_j)\big) \\
&\qquad = -\frac{2}{N^2} \cdot \frac{1}{\lambda(x_i)^2} \mathcal{L}(f)(x_i) \cdot \widetilde{\mathcal{L}}^t_{m(\underline{x})}(f)(x_i)\ .
\end{aligned}$$

Substituting in the expression of the carré du champ $\Gamma_{\overline{L}^N}$, we obtain

$$\begin{aligned}
\Gamma_{\overline{L}^N}\big(F,\, F\big)(\underline{x}) \ = &\sum_{i=1}^N \lambda(x_i) \int_E p(x_i, dy)\big(F(\underline{x}^{i,y}) - F(\underline{x})\big)^2 \\
&+ \sum_{i=1}^N \lambda(x_i) \sum_{A \in \mathcal{N}} \pi_{\underline{x}}(x_i,\, A)\big(F(\underline{x}^{A,x_i}) - F(\underline{x})\big)^2 \\
&- \frac{2}{N^2} \sum_{i=1}^N \frac{1}{\lambda(x_i)} \mathcal{L}(f)(x_i) \cdot \widetilde{\mathcal{L}}^t_{m(\underline{x})}(f)(x_i)\ . \qquad (4.19)
\end{aligned}$$

The first line in (4.19) is simply

$$\sum_{i=1}^N \lambda(x_i) \int_E p(x_i, dy)\big(F(\underline{x}^{i,y}) - F(\underline{x})\big)^2 \ = \ \frac{1}{N^2} \sum_{i=1}^N \Gamma_{\mathcal{L}}(f, f)(x_i)\ .$$

Now, considering the second line of (4.19), we can write

$$\begin{aligned}
&\lambda(x_i) \sum_{A \in \mathcal{N}} \pi_{\underline{x}}(x_i,\, A) \big(F(\underline{x}^{A,x_i}) - F(\underline{x})\big)^2 \\
&\quad = \frac{\lambda(x_i)}{N^2} \sum_{A \in \mathcal{N}} \pi_{\underline{x}}(x_i,\, A) \sum_{j,k \in A} \big(f(x_i) - f(x_j)\big) \cdot \big(f(x_i) - f(x_k)\big) \\
&\quad = \frac{\lambda(x_i)}{N^2} \Bigg(\sum_{j=1}^{N} \psi_{\underline{x}}(x_i, x_j) \big(f(x_i) - f(x_j)\big)^2 \\
&\qquad\qquad + \sum_{\substack{j,k=1 \\ k \neq j}}^{N} \Psi_{\underline{x}}(x_i;\, x_j,\, x_k) \big(f(x_i) - f(x_j)\big) \cdot \big(f(x_i) - f(x_k)\big) \Bigg) \,.
\end{aligned}$$

By hypothesis on $\psi_{\underline{x}}$, we get

$$\sum_{i=1}^{N} \frac{\lambda(x_i)}{N^2} \sum_{j=1}^{N} \psi_{\underline{x}}(x_i, x_j) \big(f(x_i) - f(x_j)\big)^2 \;=\; \frac{1}{N^2} \sum_{i=1}^{N} \Gamma_{\widetilde{\mathcal{L}}_{m(\underline{x})}}(f, f)(x_i) \,,$$

and, by hypothesis on $\Psi_{\underline{x}}$ (4.17),

$$\begin{aligned}
&\sum_{i=1}^{N} \frac{\lambda(x_i)}{N^2} \sum_{\substack{j,k=1 \\ k \neq j}}^{N} \Psi_{\underline{x}}(x_i;\, x_j,\, x_k) \big(f(x_i) - f(x_j)\big) \cdot \big(f(x_i) - f(x_k)\big) \\
&\qquad\qquad = \frac{1}{N^2} \sum_{i=1}^{N} \lambda(x_i) \cdot Q(x_i) \big(\ell_{m(\underline{x})}(f)(x_i)\big)^2 + O\Big(\frac{1}{N^2}\Big) \,.
\end{aligned}$$

Combining all together, we obtain the statement.

□

Proposition 4.2.1 and Proposition 4.2.2 ensure respectively that condition (3.3a) on the generator and condition (3.3b) on the carré du champ in Assumption 3.1.1 are satisfied for the cloning algorithm.

Roughly speaking, condition (4.17) on $\Psi_{\underline{x}}$ means that, during a cloning event for a particle i, the probability of replacing a particle $j \in \{1, \dots, N\}$ (i.e. deciding whether or not $j \in A$) provided $k_1 \in A$ or provided $k_2 \in A$ is the same for any $k_1,\, k_2 \in \{1, \dots, N\}$. Moreover, in order to satisfy condition (3.3b) on the carré du champ, condition (4.17) can be weakened by simply assuming that $\Psi_{\underline{x}}(x_i;\, x_j,\, x_k)$ has order $O\big(1/N^2\big)$, for any $i, j, k \in \{1, \dots, N\}$. The decision to use a stronger condition (4.17) for $\Psi_{\underline{x}}$ in Proposition 4.2.2 was made in order to have an explicit expression for the operator $G_\mu(f, f)$ (4.18).

Example 4.2.3. The easiest choice of a cloning generator $\overline{L}^N$ (4.13) satisfying conditions (4.14) and (4.17) is to decide whether or not to replace each particle $j \in \{1, \dots, N\}$ independently from the rest of the population with probability $\psi_{\underline{x}}(x_i, x_j)$, so that we can write

$$\pi_{\underline{x}}(x_i, A) = \prod_{j \in A} \psi_{\underline{x}}(x_i, x_j) \prod_{j \notin A} \big(1 - \psi_{\underline{x}}(x_i, x_j)\big) \ ,$$

for any $i \in \{1, \dots, N\}$ and $A \in \mathcal{N}$. Choosing $\psi_{\underline{x}}$ as in (4.14), we see that also condition (4.17) is satisfied with

$$\Psi_{\underline{x}}(x_i;\, x_j,\, x_k) \ = \ \psi_{\underline{x}}(x_i,\, x_j) \cdot \psi_{\underline{x}}(x_i,\, x_k) \ = \ \frac{Q(x_i)}{N^2}\, \omega(x_i, x_j) \cdot \omega(x_i, x_k) \ ,$$

where $Q(x) = 1/\lambda^2(x)$ and $\omega(x, y) = \widetilde{W}(y, x)$. In particular, in this case, $\ell_\mu(f) = \widetilde{\mathcal{L}}^t_\mu(f)$.

However, the construction described in Example 4.2.3 does not satisfy condition (3.3c) on the boundedness of the jumps, required for the convergence results (Theorem 3.1.2 and Theorem 3.1.3), since with this approach the number of clones per cloning event is not bounded by a constant independent of N. Indeed, in order to ensure that condition (3.3c) is satisfied, we need to assume that there exists $K > 0$ such that

$$\pi_{\underline{x}}(x, A) = 0 \text{ for all } A \in \mathcal{N} \text{ s.t. } |A| \geq K \ , \tag{4.20}$$

for any $x \in E$ and $N \in \mathbb{N}$. We further assume $N \geq K$, i.e. N is large enough so that the process (4.13) is well defined.

Example 4.2.4. A possible construction of a cloning algorithm satisfying conditions (4.14), (4.17) and (4.20) is the following. During a cloning event for the particle i, we first (randomly) decide how many particles n we want to consider according to a probability $\widetilde{\pi}(x_i, n)$; then n particles are uniformly selected from the population, so that every particle has probability n/N to be chosen; finally, every selected particle j is substituted with x_i with some probability $q(x_i, x_j)$, or kept unchanged with probability $1 - q(x_i, x_j)$. Using this construction, we have

$$\psi_{\underline{x}}(x_i, x_j) \ = \ \sum_{n=1}^{N} \widetilde{\pi}(x_i, n) \cdot \frac{n}{N} \cdot q(x_i, x_j) \ , \tag{4.21}$$

where $\widetilde{\pi}(x,n) \in [0,1]$ such that

$$\sum_{n=0}^{N} \widetilde{\pi}(x,n) = 1 \ , \quad \text{for all } x \in E \ .$$

Then, condition (4.14) is satisfied by assuming

$$M(x) := \sum_{n=1}^{N} n \cdot \widetilde{\pi}(x,n) \geq \sup_{y \in E} \frac{\widetilde{W}(y,x)}{\lambda(x)} \ , \tag{4.22}$$

independent of N, and

$$q(x,y) = \frac{\widetilde{W}(y,x)}{M(x) \cdot \lambda(x)} \in [0,1] \ . \tag{4.23}$$

Finally, in order to satisfy (4.20), we assume that

$$\widetilde{\pi}(x,n) = 0 \quad \text{for all } n \geq K \text{ and } x \in E \ . \tag{4.24}$$

One common choice for $\widetilde{\pi}(x,n)$ is

$$\widetilde{\pi}(x,n) = \begin{cases} M(x) - \lfloor M(x) \rfloor & n = \lfloor M(x) \rfloor + 1, \\ \lfloor M(x) \rfloor + 1 - M(x) & n = \lfloor M(x) \rfloor, \\ 0 & \text{otherwise.} \end{cases} \tag{4.25}$$

This corresponds to a binary distribution on the two integers nearest to the prescribed mean $M(x)$, and minimises the second moment Q of the distribution for a given mean. Note that if $M(x)$ is an integer, $\widetilde{\pi}(x,n) = \delta_{n,M(x)}$ concentrates, which includes the case $M(x) = 0$.

Under this construction, condition (4.17) holds with

$$Q(x) = \frac{\sum_{n=0}^{K} \widetilde{\pi}(x,n) \cdot n(n-1)}{M^2(x) \cdot \lambda^2(x)} \quad \text{and} \quad \omega(x,y) = \widetilde{W}(y,x) \ , \tag{4.26}$$

where $K > 0$ is given by (4.24), so that $Q(x)$ is independent of N and condition (3.3c) on bounded jumps is satisfied.

Remark. In case $\widetilde{W}(y,x) = \widetilde{W}(y,x)$ is independent of $y \in E$, i.e. $\widetilde{W}(y,x) = h(x)$ for some $h \in \mathcal{C}_b(E)$ and for any $x, y \in E$, then conditions (4.22)-(4.23) are satisfied by taking $q(x,y) \equiv 1$ and $M(x) = h(x)/\lambda(x)$. The independence of $\widetilde{W}(x,y)$ from the first variable $x \in E$ occurs for instance with the McKean model given by (2.21a)

with $c \leq \inf \mathcal{V}$. This particular case is illustrated by Angeli et al. (2019, Section 3.3).

Under the initial condition (3.3d) and assuming (4.20), Proposition 4.2.1 and Proposition 4.2.2 show in particular that Assumption 3.1.1 is satisfied for cloning algorithms, hence Theorem 3.1.2 holds and provides bias and L^p error bounds.

Proposition 4.2.5 (CLT). *Let μ_t^N be the empirical measure of an N-particle cloning process satisfying conditions (3.3d), (4.14), (4.17) and (4.20). Furthermore, assume that*

$$\lim_{z \to x} \|\widetilde{W}(z, \cdot) - \widetilde{W}(x, \cdot)\| = 0 \ , \quad \lim_{z \to x} \|\widetilde{W}(\cdot, z) - \widetilde{W}(\cdot, x)\| = 0 \ , \tag{4.27}$$

and that

$$\lim_{z \to x} \|\omega(z, \cdot) - \omega(x, \cdot)\| = 0 \ . \tag{4.28}$$

Then, under the initial condition (3.3d), for any fixed $T > 0$, the sequence

$$V_T^N(f) := \sqrt{N}\big(\mu_T^N(f) - \mu_T(f)\big) \ ,$$

converges in law as $N \to \infty$ to a centered Gaussian random variable $V_T(f)$ with variance given by

$$\begin{aligned} &\mathbb{E}\big[V_T(f)^2\big] \\ &\quad = \mu_T\big((f - \mu_T(f))^2\big) + \int_0^T \mu_s\Big((\Theta_{s,T}\overline{f})^2 \cdot \big(\mathcal{V} - \mu_s(\mathcal{V})\big) + \Gamma_{\widetilde{\mathcal{L}}_{\mu_s}}(\Theta_{s,T}\overline{f}, \Theta_{s,T}\overline{f}) \\ &\qquad + \lambda Q\left(\ell_{\mu_s}(\Theta_{s,T}\overline{f})\right)^2 - \frac{2}{\lambda}\mathcal{L}(\Theta_{s,T}\overline{f}) \cdot \widetilde{\mathcal{L}}_{\mu_s}^t(\Theta_{s,T}\overline{f})\Big)\, ds \ . \end{aligned} \tag{4.29}$$

Proof. For any $f \in \mathcal{C}_b(E)$, the operator $G_\mu(f, f)$ (4.18) is in the form (3.32), with

$$\begin{aligned} g_0(x) &= \Gamma_{\mathcal{L}}(f, f)(x) \ , \\ g_1(y, x) &= \widetilde{W}(x, y)\big(f(y) - f(x)\big)^2 - \frac{2}{\lambda(x)}\mathcal{L}(f)(x)\,\widetilde{W}(y, x)\big(f(y) - f(x)\big) \ , \\ g_2(y, x) &= \sqrt{\lambda(x) \cdot Q(x)} \cdot \omega(x, y)\big(f(y) - f(x)\big) \ , \end{aligned}$$

independent of the distribution $\mu \in \mathcal{P}(E)$.

Under conditions (4.27) and (4.28), the hypotheses in Lemma 3.3.2 are satisfied and then Theorem 3.1.3 holds and provides a central limit theorem for such algorithms.

Similarly to Proposition 4.1.1, the conclusion follows by (4.10).

□

Remark. Observe that condition (4.27) is satisfied for all three McKean models (2.21a)-(2.21c), as discussed in the remark below Theorem 4.1.1. Furthermore, this implies that condition (4.28) holds for cloning algorithms satisfying (4.22)-(4.23), since in this case $\omega(x,y) = \widetilde{W}(y,x)$.

4.2.2 Some remarks on the asymptotic variance

As for mean field particle approximations (Section 4.1.2), in order to improve the accuracy of cloning algorithms, we need to reduce the asymptotic variance $\mathbb{E}\big[V_T(f)^2\big]$, which can be written in the form (4.29), provided the hypotheses in Proposition 4.2.5 are satisfied. However, $\mathbb{E}\big[V_T(f)^2\big]$ cannot be estimated a priori and, moreover, it is not possible to say a priori whether a mean field particle approximation is better than a cloning algorithm or not. Indeed, comparing the central limit theorem for mean field particle approximations (Proposition 4.1.1) with the one for cloning algorithms (Proposition 4.2.5), we can see that, given a McKean model $\overline{\mathcal{L}}_{\mu_T}$, the difference of the asymptotic variances of V_T^{MF} and V_T^{clon} is given by

$$\mathbb{E}\big[V_T^{clon}(f)^2\big] - \mathbb{E}\big[V_T^{MF}(f)^2\big] = \int_0^T \mu_s\Big(\lambda\, Q\left(\ell_{\mu_s}(\Theta_{s,T}\overline{f})\right)^2 - \frac{2}{\lambda}\,\mathcal{L}(\Theta_{s,T}\overline{f})\cdot\widetilde{\mathcal{L}}^t_{\mu_s}(\Theta_{s,T}\overline{f})\Big)\,ds\ .$$

Whereas the term $\lambda\, Q\left(\ell_{\mu_s}(\Theta_{s,T}\overline{f})\right)^2$ is always non-negative (and vanishes when there is at most one clone, i.e. when $\Psi_{\underline{x}} \equiv 0$), meaning that the dependence between clones brings an extra error, the sign of the term $\mathcal{L}(\Theta_{s,T}\overline{f})\cdot\widetilde{\mathcal{L}}^t_{\mu_s}(\Theta_{s,T}\overline{f})$ is not known a priori and also depends on time, because of the propagator $\Theta_{s,T}$. However, observe that this term is independent of the particular choice of cloning algorithm, but depends on the McKean model. In particular, given a McKean model, we can optimise the construction of the cloning algorithm.

Indeed, consider two sequences of cloning processes $\big(\overline{L}^{N,(1)}\big)_{N\in\mathbb{N}}$ and $\big(\overline{L}^{N,(2)}\big)_{N\in\mathbb{N}}$ associated to a given McKean model $\big(\overline{\mathcal{L}}_\mu\big)_{\mu\in\mathcal{P}(E)}$, such that the hypotheses for the Central Limit Theorem for cloning processes (Proposition 4.2.5) hold and satisfy the construction given in Example 4.2.4 with $Q^{(1)}$ and $\omega^{(1)}$ (resp. $Q^{(2)}$ and $\omega^{(2)}$) in the form (4.26), i.e.

$$Q(x) = \frac{\sum_{n=0}^K \widetilde{\pi}(x,n)\cdot n(n-1)}{M(x)^2\cdot\lambda(x)^2} \quad \text{and} \quad \omega(x,y) = \widetilde{W}(y,x)\ ,$$

for some $K > 0$ independent of $N \in \mathbb{N}$, and hence $\ell_\mu(f) = \widetilde{\mathcal{L}}^t_\mu(f)$, for any $f \in \mathcal{C}_b(E)$. By the Central Limit Theorem for cloning processes (Proposition 4.2.5), we have

that

$$\mathbb{E}\big[V_T^{(1)}(f)^2\big] - \mathbb{E}\big[V_T^{(2)}(f)^2\big] = \int_0^T \mu_s\Big(\lambda\, Q^{(1)}\big(\ell_{\mu_s}^{(1)}(\Theta_{s,T}\overline{f})\big)^2 \;-\; \lambda\, Q^{(2)}\big(\ell_{\mu_s}^{(2)}(\Theta_{s,T}\overline{f})\big)^2\Big)\, ds\,.$$

By construction of $Q^{(1)}$ and $\ell^{(1)}$ (resp. $Q^{(2)}$ and $\ell^{(2)}$), we obtain

$$\mathbb{E}\big[V_T^{(1)}(f)^2\big] - \mathbb{E}\big[V_T^{(2)}(f)^2\big] \;=\; \int_0^T \mu_s\Big(\lambda\,(Q^{(1)} - Q^{(2)})\cdot\big(\widetilde{\mathcal{L}}_{\mu_s}^t(\Theta_{s,T}\overline{f})\big)^2\Big)\, ds\;.$$

Therefore, it is easy to see that the asymptotic variance is minimised by minimising $Q(x)$, subject to $M(x) \geq \sup_{y\in E} \widetilde{W}(y,x)/\lambda(x)$, by (4.22). Fixing $M(x)$, we see that the optimal $Q(x)$ is obtained by minimising the second moment $\sum \widetilde{\pi}(x,n)\cdot n(n-1)$, and this is achieved by the common choice (4.25). In particular, with this choice, in case $M(x) \leq 1$, we have $\widetilde{\pi}(x,n) = 0$ for any $n > 1$ and thus $Q(x) = 0$.

4.2.3 The classical cloning algorithm

We conclude the section with a brief description of a variant of the cloning algorithm proposed by Lecomte and Tailleur (2007) and commonly used in the theoretical physics literature. This procedure is constructed from the McKean model $\overline{\mathcal{L}}_\mu$ (2.16) with selection rates $\widetilde{W}_c(x,y) = \big(\mathcal{V}(x)-c\big)^- + \big(\mathcal{V}(y)-c\big)^+$ as in (2.21a), and we denote the associated McKean generator by

$$\overline{\mathcal{L}}_\mu(f)(x) := \mathcal{L}(f)(x) + \int_E \widetilde{W}_c(x,y)\big(f(y)-f(x)\big)\mu(dy)\;. \tag{4.30}$$

The infinitesimal description of this cloning algorithm as a continuous-time Markov process on the state space E^N is given by the generator

$$\begin{aligned}\overline{L}_c^N(F)(\underline{x}) \;=\; &\sum_{i=1}^N \lambda(x_i)\int_E p(x_i,\,dy)\sum_{A\in\mathcal{N}} \pi_{\underline{x}}(x_i,\,A)\cdot\big(F(\underline{x}^{A,x_i;\,i,y}) - F(\underline{x})\big)\\ &+\sum_{i=1}^N \big(\mathcal{V}(x_i)-c\big)^- \frac{1}{N}\sum_{j=1}^N \big(F(\underline{x}^{i,x_j}) - F(\underline{x})\big),\end{aligned} \tag{4.31}$$

for any $F \in \mathcal{C}_b(E^N)$ and $\underline{x} \in E^N$, where the probabilities $\pi_{\underline{x}}(x_i, A)$ are such that

$$\psi_{\underline{x}}(x_i, x_j) \;=\; \sum_{A|j\in A} \pi_{\underline{x}}(x_i, A) \;=\; \frac{\big(\mathcal{V}(x_i) - c\big)^+}{N\,\lambda(x_i)} \;,$$

so to satisfy condition (4.14), which ensures the correct expected behaviour of mean-field observables. With this version of cloning algorithm, if $\mathcal{V}(x_i) > c$, a non-empty set A of particles is chosen at random from the ensemble with probability $\pi_{\underline{x}}(x_i,\, A)$ and every particle $j \in A$ is replaced by a clone of i, before particle i mutates to a new state $y \in E$. If $\mathcal{V}(x_i) \leq c$ we set $\pi_{x_i}(A) = \delta_{A,\emptyset}$, so that no cloning occurs. The killing part in the second line runs independently.

A common choice for $\pi_{\underline{x}}(x_i, A)$ in the theoretical physics literature is given by the construction in Example 4.2.4, with $\widetilde{\pi}(x, n)$ given by the common choice (4.25), where

$$M(x) = \frac{\big(\mathcal{V}(x) - c\big)^+}{\lambda(x)}$$

gives the average number of clones and $q(x, y) \equiv 1$. With this construction, the probability of replacing a particle j during a cloning event for a particle i is given by

$$\psi_{\underline{x}}(x_i, x_j) \;=\; \sum_{n=1}^{N} \frac{n}{N}\,\widetilde{\pi}(x_i, n) \;=\; \frac{M(x)}{N} \;=\; \frac{\big(\mathcal{V}(x) - c\big)^+}{N\,\lambda(x)} \;,$$

hence condition (4.14) on $\psi_{\underline{x}}$ is satisfied. Furthermore, under this construction, condition (4.17) holds with

$$Q(x) = \frac{\sum_{n=K}^{K+1} \widetilde{\pi}(x, n)\cdot n(n-1)}{\Big(\big(\mathcal{V}(x) - c\big)^+\Big)^2} \;, \quad \text{with } K = \lfloor M(x) \rfloor \;,$$

and $\omega(x, y) = \big(\mathcal{V}(x) - c\big)^+$.

We will consider this variant of the cloning algorithm in Section 5.4, where we will construct estimators based on the so-called 'cloning factor'.

4.3 The resampling algorithm

We present here a third interacting particle approximation, the resampling algorithm, as an example of a particle approximation that does not satisfy Assumption 3.1.1, and we show indeed that the convergence results proved in Chapter 3 do not hold for the considered class of resampling algorithms. The results shown

in this section help to better understand the relevance of Assumption 3.1.1, and in particular of condition 3.3b on the carré du champ, in the study of the convergence of an interacting particle approximation.

The (continuous-time) resampling algorithm presented here consists in running in parallel mutation and selection (or resampling) events, independently from each other. The mutation dynamics coincides with the mutation for mean field particle systems, i.e. (4.2), whereas, mimicking the basic idea for discrete-time resampling procedures (see, for example, Douc et al., 2005; Hol et al., 2006; Doucet and Johansen, 2011), the resampling dynamics gathers selection events together and can be described as follows:

- the resampling events have escape rate $\lambda^{\Delta}(\underline{x})$;
- at each resampling event, a set B of particles is chosen with probability $p^{\Delta}(\underline{x};\ B)$, so that
$$\sum_{B\in\mathcal{N}} p^{\Delta}(\underline{x};\ B) = 1\ ;$$
- every element $i \in B$ is killed and replaced by a clone of a particle j with probability $q^{\Delta}_{m(\underline{x})}(x_i, x_j)$, so that
$$\sum_{j=1}^{N} q^{\Delta}_{m(\underline{x})}(x_i, x_j) = 1\ ,$$
with $j = 1, \ldots, N$.

Remark. The quantities λ^{Δ}, p^{Δ} and $q^{\Delta}_{m(\cdot)}$ possibly depend on N, but we omit writing explicitly the dependence with N, to avoid notation overload.

With this construction, the infinitesimal generator of a resampling process can be written as
$$L^N \ = \ L^{mut,N} \ + \ L^{\Delta,\,N}, \tag{4.32}$$
with $L^{mut,N} := \sum_i \mathcal{L}^{(i)}$ and
$$L^{\Delta,\,N}(F)(\underline{x}) \ = \ \lambda^{\Delta}(\underline{x}) \sum_{B\in\mathcal{N}} p^{\Delta}(\underline{x};\ B)\ \cdot \\ \sum_{\underline{\delta}\in\{1,\ldots,N\}^{|B|}} \left(\prod_{k=1}^{|B|} q^{\Delta}_{m(\underline{x})}(x_{b_k},\ x_{\delta_k}) \right) \cdot \left(F\left(\underline{x}^{B,\underline{\delta}}\right) - F(\underline{x}) \right)\ . \tag{4.33}$$
where $b_1 < \cdots < b_{|B|}$ denote the elements of B and $\underline{x}^{B,\underline{\delta}}$ denotes the vector

$(z_1, \dots, z_N) \in E^N$ with

$$z_i := \begin{cases} x_i & i \notin B, \\ x_{\delta_k} & i = b_k,\ k \in \{1, \dots, |B|\}. \end{cases}$$

To simplify the presentation, we limit ourselves to consider resampling processes in which the resampling escape rate $\lambda^\Delta(\underline{x}) = \lambda^\Delta$ is independent of the state $\underline{x}$ of the system and the particles to be replaced are selected independently from the whole populations, so that

$$\sum_{B|i,k\in B} p^\Delta(\underline{x};\, B) \;=\; P^\Delta(\underline{x};\, i) \cdot P^\Delta(\underline{x};\, k), \tag{4.34}$$

for any $i \neq k$, where

$$P^\Delta(\underline{x};\ i) := \sum_{B|i\in B} p^\Delta(\underline{x};\ B)$$

is the probability of i being chosen to be replaced. However, this assumption is not necessary and different constructions are possible.

Assumption 4.3.1. Given a family of McKean generators $\big(\overline{\mathcal{L}}_\mu\big)_{\mu\in\mathcal{P}(E)}$ (2.16)-(2.18), we assume that the sequence of interacting particle approximations with resampling generators L^N (4.32)-(4.33), $N \in \mathbb{N}$, satisfies

$$\begin{aligned} &a) \quad \lambda^\Delta \geq \sup_{i\in\{1,\dots,N\}} m(\underline{x})\big(\widetilde{W}(x_i,\, \cdot\,)\big)\,, \\ &b) \quad P^\Delta(\underline{x};\ i) = \frac{m(\underline{x})\big(\widetilde{W}(x_i,\, \cdot\,)\big)}{\lambda^\Delta}\,, \\ &c) \quad q^\Delta_{m(\underline{x})}(x_i, x_j) = \frac{\widetilde{W}(x_i,\, x_j)}{N \cdot m(\underline{x})\big(\widetilde{W}(x_i,\, \cdot\,)\big)}\,, \end{aligned}$$

for any configuration $\underline{x}$ and $i,\, j \in \{1, \dots, N\}$.

Example 4.3.2. Let $\overline{\mathcal{L}}_\mu$ be the McKean model given by (2.21a) with $c = \sup_x \mathcal{V}(x)$. One possible choice of parameters for a resampling algorithm associated to $\overline{\mathcal{L}}_\mu$ is the following:

$$\begin{aligned} \lambda^\Delta \;&\equiv\; \sup_x \mathcal{V}(x) \;-\; \inf_x \mathcal{V}(x)\,, \\ P^\Delta(\underline{x};\, i) \;&=\; \frac{\sup_x \mathcal{V}(x) \;-\; \mathcal{V}(x_i)}{\sup_x \mathcal{V}(x) \;-\; \inf_x \mathcal{V}(x)}\,, \\ q^\Delta_{m(\underline{x})}(x_i,\, x_j) \;&=\; \frac{1}{N}\,. \end{aligned}$$

With this choice, the rate of resampling events has order 1 and, during each resampling event, the size of the set B of particles that get replaced is of order N.

Example 4.3.3. As for the previous example, we consider the McKean model given by (2.21a) with $c = \sup_x \mathcal{V}(x)$. A second possible choice of parameters for an N-particle resampling algorithm associated to $\overline{\mathcal{L}}_\mu$ is the following:

$$\begin{aligned} \lambda^\Delta &\equiv N \cdot \big(\sup_x \mathcal{V}(x) \,-\, \inf_x \mathcal{V}(x)\big)\,, \\ P^\Delta(\underline{x};\, i) &= \frac{\sup_x \mathcal{V}(x) \,-\, \mathcal{V}(x_i)}{N\big(\sup_x \mathcal{V}(x) \,-\, \inf_x \mathcal{V}(x)\big)}\,, \\ q^\Delta_{m(\underline{x})}(x_i,\, x_j) &= \frac{1}{N}\,. \end{aligned}$$

However, this second choice is highly expensive from a computational point of view, since the rate of resampling events has order N, in particular the number of resampling events in a given time interval goes to infinity as N goes to infinity. Furthermore, at every resampling event, each particle gets replaced with only a small probability of order $1/N$. Hence, the probability $p^\Delta(x, B)$ of resampling a set B of particles is of the order $1/N^{|B|}$, and therefore the size of resampling events is typically of order 1. This is in line with mean field particle approximations, where the overall selection rate is of order N and the size of selection events is 1. However, we can easily see that the computational cost of this version of resampling is N times higher than the cost of the mean field particle approximation (associated to the same McKean model given by (2.21a) with $c = \sup_x \mathcal{V}(x)$), since at every resampling event the algorithm runs through the whole population and checks each particle independently to decide whether or not to replace it. Therefore, it would be impractical to implement this version of the resampling algorithm.

Proposition 4.3.4. *Given a family of McKean generators* $\big(\overline{\mathcal{L}}_\mu\big)_{\mu\in\mathcal{P}(E)}$ (2.16)-(2.18), *with selection part denoted by* $\widetilde{\mathcal{L}}_\mu$, *consider a sequence of resampling processes with generators* L^N (4.32)-(4.33), $N \in \mathbb{N}$, *and satisfying Assumption 4.3.1. Then, for any function* $F \in \mathcal{C}_b(E^N)$ *in the form* $F(\underline{x}) = m(\underline{x})(f)$, $f \in \mathcal{C}_b(E)$,

$$L^N(F)(\underline{x}) \;=\; m(\underline{x})\big(\overline{\mathcal{L}}_{m(\underline{x})}(f)\big)\;, \tag{4.35}$$

and

$$\Gamma_{L^N}(F,F)(\underline{x}) \;=\; \frac{1}{\lambda^\Delta}\Big(m(\underline{x})\big(\widetilde{\mathcal{L}}_{m(\cdot)}(f)\big)\Big)^2 \;+\; \frac{1}{N}\,m(\underline{x})\big(G_{m(\cdot)}(f,f)\big)\;, \tag{4.36}$$

where $(G_\mu)_{\mu\in\mathcal{P}(E)}$ *is a family of bilinear operators given by*

$$G_\mu(f,f) := \Gamma_{\overline{\mathcal{L}}_\mu}(f,f) \;-\; \frac{1}{\lambda^\Delta}\big(\widetilde{\mathcal{L}}_\mu(f)\big)^2,$$

for any $F\in\mathcal{C}_b(E^N)$ *in the form* $F(\underline{x}) = m(\underline{x})(f)$, $f\in\mathcal{C}_b(E)$.

Proof. For proving Equation (4.35), it is enough to observe that, for the considered class of functions, given a set $B\in\mathcal{N}$, we can write

$$\begin{aligned}
&\sum_{\underline{\delta}\in\{1,\dots,N\}^{|B|}}\Big(\prod_{k=1}^{|B|} q^\Delta_{m(\underline{x})}(x_{b_k},\,x_{\delta_k})\Big)\cdot\Big(F\big(\underline{x}^{B,\underline{\delta}}\big)-F(\underline{x})\Big)\\
&\qquad=\frac{1}{N}\sum_{\underline{\delta}\in\{1,\dots,N\}^{|B|}}\Big(\prod_{k=1}^{|B|} q^\Delta_{m(\underline{x})}(x_{b_k},\,x_{\delta_k})\Big)\cdot\sum_{k=1}^{|B|}\big(f(x_{\delta_k})-f(x_{b_k})\big)\\
&\qquad=\frac{1}{N}\sum_{\underline{\delta}\in\{1,\dots,N\}^{|B|}}\sum_{k=1}^{|B|} q^\Delta_{m(\underline{x})}(x_{b_k},\,x_{\delta_k})\cdot\big(f(x_{\delta_k})-f(x_{b_k})\big)\\
&\qquad=\frac{1}{N}\sum_{i\in B}\sum_{j=1}^{N} q^\Delta_{m(\underline{x})}(x_i,\,x_j)\cdot\big(f(x_j)-f(x_i)\big)\;. \qquad (4.37)
\end{aligned}$$

Thus, the infinitesimal generator can be rewritten as

$$\begin{aligned}
L^{\Delta,N}(F)(\underline{x}) &= \lambda^\Delta\sum_{B\in\mathcal{N}} p^\Delta(\underline{x};\,B)\sum_{i\in B}\sum_{j=1}^{N} q^\Delta_{m(\underline{x})}(x_i,\,x_j)\,\frac{f(x_j)-f(x_i)}{N}\\
&= \lambda^\Delta\sum_{i=1}^{N} P^\Delta(\underline{x};\,i)\sum_{j=1}^{N} q^\Delta_{m(\underline{x})}(x_i,\,x_j)\,\frac{f(x_j)-f(x_i)}{N}\;.
\end{aligned}$$

Substituting the quantities λ^Δ, P^Δ and $q^\Delta_{m(\underline{x})}$ as in Assumption 4.3.1, we obtain

$$L^{\Delta,N}(F)(\underline{x}) \;=\; m(\underline{x})\big(\widetilde{\mathcal{L}}_{m(\underline{x})}(f)\big)\;.$$

Including the mutation dynamics, we then obtain (4.35).

To prove Equation 4.36, first we recall that the carré du champ $\Gamma_{L^{\Delta,N}}$ asso-

ciated to the resampling generator $L^{\Delta,N}$ (4.33) can be written as

$$\Gamma_{L^{\Delta,N}}(F,F)(\underline{x}) = \lambda^{\Delta} \sum_{B\in\mathcal{N}} p^{\Delta}(\underline{x};B) \sum_{\underline{\delta}\in\{1,\dots,N\}^{|B|}} \left(\prod_{k=1}^{|B|} q^{\Delta}_{m(\underline{x})}(x_{b_k}, x_{\delta_k}) \right) \cdot \left(F\big(\underline{x}^{B,\underline{\delta}}\big) - F(\underline{x}) \right)^2 . \tag{4.38}$$

Note that for functions in the form $F(\underline{x}) = m(\underline{x})(f)$, $f \in \mathcal{C}_b(E)$,

$$\big(F\big(\underline{x}^{B,\underline{\delta}}\big) - F(\underline{x})\big)^2 = \frac{1}{N^2} \sum_{k=1}^{|B|} \big(f(x_{\delta_k}) - f(x_{b_k})\big)^2 + \frac{1}{N^2} \sum_{\substack{k,l\in B \\ k\neq l}} \big(f(x_{\delta_k}) - f(x_{b_k})\big) \cdot \big(f(x_{\delta_l}) - f(x_{b_l})\big) .$$

Similarly to (4.37), for any fixed set $B \in \mathcal{N}$ we can write

$$\sum_{\underline{\delta}\in\{1,\dots,N\}^{|B|}} \left(\prod_{k=1}^{|B|} q^{\Delta}_{m(\underline{x})}(x_{b_k}, x_{\delta_k}) \right) \cdot \sum_{k=1}^{|B|} \big(f(x_{\delta_k}) - f(x_{b_k})\big)^2 = \sum_{i\in B} \sum_{j=1}^{N} q^{\Delta}_{m(\underline{x})}(x_i,\, x_j) \cdot \big(f(x_j) - f(x_i)\big)^2 ,$$

and

$$\sum_{\underline{\delta}\in\{1,\dots,N\}^{|B|}} \left(\prod_{k=1}^{|B|} q^{\Delta}_{m(\underline{x})}(x_{b_k}, x_{\delta_k}) \right) \cdot \sum_{\substack{k,l\in B \\ k\neq l}} \big(f(x_{\delta_k}) - f(x_{b_k})\big) \cdot \big(f(x_{\delta_l}) - f(x_{b_l})\big) = \sum_{\substack{i,k\in B \\ i\neq k}} \sum_{j,l=1}^{N} q^{\Delta}_{m(\underline{x})}(x_i, x_j)\, q^{\Delta}_{m(\underline{x})}(x_k, x_l) \cdot \big(f(x_j) - f(x_i)\big)\big(f(x_l) - f(x_j)\big) .$$

Substituting in (4.38) and using condition (4.34), we get

$$\Gamma_{L^{\Delta,N}}(F,F)(\underline{x}) = \frac{\lambda^{\Delta}}{N^2} \sum_{i=1}^{N} P^{\Delta}(\underline{x};i) \sum_{j=1}^{N} q^{\Delta}_{m(\underline{x})}(x_i,\, x_j) \cdot \big(f(x_j) - f(x_i)\big)^2 + \frac{\lambda^{\Delta}}{N^2} \sum_{\substack{i,k=1 \\ i\neq k}}^{N} P^{\Delta}(\underline{x};i) \cdot \cdot P^{\Delta}(\underline{x};j) \sum_{j,l=1}^{N} q^{\Delta}_{m(\underline{x})}(x_i, x_j)\, q^{\Delta}_{m(\underline{x})}(x_k, x_l) \cdot \big(f(x_j) - f(x_i)\big)\big(f(x_l) - f(x_j)\big) .$$

Applying the substitutions given in Assumption 4.3.1, we obtain

$$\Gamma_{L^{\Delta,N}}(F,F)(\underline{x}) = \frac{1}{N} m(\underline{x})\Big(\Gamma_{\widetilde{\mathcal{L}}_{m(\underline{x})}}(f)\Big) + \frac{1}{N^2 \cdot \lambda^\Delta} \sum_{\substack{i,k=1 \\ i \neq k}}^{N} \widetilde{\mathcal{L}}_{m(\underline{x})}(f)(x_i) \cdot \widetilde{\mathcal{L}}_{m(\underline{x})}(f)(x_k) \ .$$

The conclusion follows by splitting the summation as

$$\frac{1}{\lambda^\Delta \cdot N^2} \sum_{\substack{i,k \\ i \neq k}} \widetilde{\mathcal{L}}_{m(\underline{x})}(f)(x_i) \cdot \widetilde{\mathcal{L}}_{m(\underline{x})}(f)(x_k) = \frac{1}{\lambda^\Delta}\Big(m(\underline{x})\big(\widetilde{\mathcal{L}}_{m(\underline{x})}(f)\big)\Big)^2 - \frac{1}{\lambda^\Delta \cdot N} m(\underline{x})\Big(\big(\widetilde{\mathcal{L}}_{m(\underline{x})}(f)\big)^2\Big) \ .$$

Including the mutation dynamics, which is independent from the resampling, we obtain (4.36). □

Proposition 4.3.4 shows that the family of resampling processes (L^N) satisfies condition (3.3a) on the generator, but the carré du champ differs from (3.3b) by the extra term

$$\frac{1}{\lambda^\Delta}\Big(m(\underline{x})\big(\widetilde{\mathcal{L}}_{m(\underline{x})}(f)\big)\Big)^2 \ . \tag{4.39}$$

In case $\lambda^\Delta = O(N)$, as for the Example 4.3.3, condition (3.3b) is satisfied, however this procedure would be of little practicality in terms of applications, as discussed in Example 4.3.3. On the other hand, if we consider resampling algorithms in which the resampling rate λ^Δ is a constant independent of the population size N (as seen in Example 4.3.2, for instance), (4.39) does not vanish as N goes to infinity. Therefore we expect μ_t^N not to concentrate on μ_t in the limit, and thus L^p errors will not converge with rate $1/\sqrt{N}$, as opposed to mean field approximations and cloning algorithms, as illustrated in the following result.

Proposition 4.3.5. *Consider a sequence of resampling processes satisfying Assumption 4.3.1, with λ^Δ constant independent of N, with empirical distribution μ_t^N, and with initial configuration satisfying (3.3d). Then, for any $T \geq 0$, $p > 2$ and for any $f \in \mathcal{C}_b(E)$ such that $\mu_s(\mathcal{V} \cdot \Theta_{s,T} f) \neq \mu_s(\mathcal{V}) \cdot \mu_T(f)$, for some $s \in [0,T]$, there does NOT exist any constant $c_p > 0$ independent of N such that*

$$\mathbb{E}\left[\big(\mu_T^N(f) - \mu_T(f)\big)^p\right]^{1/p} \leq \frac{c_p \|f\|}{N^{1/2}} \ , \tag{4.40}$$

for any $N \in \mathbb{N}$ large enough.

Remark. Observe that the condition $\mu_s(\mathcal{V}\cdot\Theta_{s,T}f) \neq \mu_s(\mathcal{V})\cdot\mu_T(f)$ for some $s\in[0,T]$ is always ensured if $\mu_T(\mathcal{V}\cdot f)\neq\mu_T(\mathcal{V})\cdot\mu_T(f)$, by taking $s=T$. This occurs for example when $f(x)=\mathcal{V}(x)$ and $\mathcal{V}$ is not constant a.e. on the domain of μ_T.

Proof. Given $f\in\mathcal{C}_b(E)$, suppose that the bound (4.40) holds for any $T\geq 0$, with c_p independent of T. Then, in the same way as for (3.6), $\mu_t^N(f)$ converges a.s. to $\mu_t(f)$ as $N\to\infty$, for any $t\geq 0$ and $f\in\mathcal{C}_b(E)$. Similarly to the argument used in Section 3.2.1, we can write

$$\mu_T^N(f)-\mu_T(f) \;=\; \mu_0^N(\Theta_{0,T}\overline{f}) + \mathcal{M}_T^N(\Theta_{\cdot,T}\overline{f}) + \int_0^T \mu_s^N\big(\Theta_{s,T}\overline{f}\big)\cdot\big(\mu_s(\mathcal{V})-\mu_s^N(\mathcal{V})\big)\,ds\;, \tag{4.41}$$

where the martingale $\mathcal{M}_T^N\big(\Theta_{\cdot,T}\overline{f}\big)$ has predictable quadratic variation given by

$$\big\langle\mathcal{M}^N\big(\Theta_{\cdot,T}\overline{f}\big)\big\rangle_t \;=\; \int_0^t \frac{1}{N}\,\mu_s^N\big(G_{\mu_s^N}(\Theta_{s,T}\overline{f},\,\Theta_{s,T}\overline{f})\big) \;+\; \frac{1}{\lambda^\Delta}\Big(\mu_s^N\big(\widetilde{\mathcal{L}}_{\mu_s^N}(\Theta_{s,T}\overline{f})\big)\Big)^2\,ds\;,$$

by using the formulation of the carré du champ (4.36) for resampling algorithms. Under the assumption that the bound (4.40) holds, we can apply Lemma 3.3.1 and obtain

$$\lim_{N\to\infty}\int_0^T \Big|\mu_s^N(\Theta_{s,T}\overline{f})\cdot\big(\mu_s(\mathcal{V})-\mu_s^N(\mathcal{V})\big)\Big|\,ds \;=\; 0\;,$$

in probability. Moreover, as $N\to\infty$, $\mu_0^N(\Theta_{0,T}\overline{f})$ converges to $\mu_0(\Theta_{0,T}\overline{f})=0$ in probability by the initial condition (3.3d), and also $\mu_T^N(f)-\mu_T(f)$ converges to 0 in probability, by assumption. Thus, using the identity (4.41), we see that the martingale $\mathcal{M}_T^N(\Theta_{\cdot,T}f)$ converges in law to 0 as $N\to\infty$ and, in particular,

$$\lim_{N\to\infty}\mathbb{E}\Big[\big\langle\mathcal{M}^N\big(\Theta_{\cdot,T}\overline{f}\big)\big\rangle_T\Big] \;=\; 0\;.$$

On the other hand, by Fatou's Lemma,

$$\begin{aligned}\lim_{N\to\infty}\mathbb{E}\Big[\big\langle\mathcal{M}^N\big(\Theta_{\cdot,T}\overline{f}\big)\big\rangle_T\Big] \;&=\; \lim_{N\to\infty}\mathbb{E}\Big[\int_0^T \frac{1}{\lambda^\Delta}\Big(\mu_s^N\big(\widetilde{\mathcal{L}}_{\mu_s^N}(\Theta_{s,T}\overline{f})\big)\Big)^2\,ds\Big]\\ &\geq\; \mathbb{E}\Big[\int_0^T \liminf_{N\to\infty}\frac{1}{\lambda^\Delta}\Big(\mu_s^N\big(\widetilde{\mathcal{L}}_{\mu_s^N}(\Theta_{s,T}\overline{f})\big)\Big)^2\,ds\Big]\;,\end{aligned}$$

and, by the characterisation (2.15) of McKean models, the quantity

$$\mu_s^N\big(\widetilde{\mathcal{L}}_{\mu_s^N}(\Theta_{s,T}\overline{f})\big) \;=\; \mu_s^N(\mathcal{V}\cdot\Theta_{s,T}\overline{f}) \;-\; \mu_s^N(\mathcal{V})\cdot\mu_s^N(\Theta_{s,T}\overline{f})$$

converges a.s. to

$$\mu_s(\mathcal{V} \cdot \Theta_{s,T}\overline{f}) \; - \; \mu_s(\mathcal{V}) \cdot \mu_s(\Theta_{s,T}\overline{f}) \; ,$$

as $N \to \infty$, by assumption. Thus,

$$\begin{aligned} \lim_{N\to\infty} \mathbb{E}\Big[\big\langle \mathcal{M}^N(\Theta_{\cdot,T}\overline{f})\big\rangle_T\Big] \; &\geq \; \frac{1}{\lambda^\Delta} \int_0^T \Big(\mu_s(\mathcal{V} \cdot \Theta_{s,T}\overline{f}) \; - \; \mu_s(\mathcal{V}) \cdot \mu_s(\Theta_{s,T}\overline{f})\Big)^2 \, ds \\ &= \; \frac{1}{\lambda^\Delta} \int_0^T \Big(\mu_s(\mathcal{V} \cdot \Theta_{s,T}\overline{f})\Big)^2 \, ds \; , \end{aligned}$$

since $\mu_s(\Theta_{s,T}\overline{f}) = \mu_T(\overline{f}) = 0$, where the right-hand side is zero only if $\mu_s(\mathcal{V}\cdot\Theta_{s,T}\overline{f}) = 0$ for almost every $s \geq 0$. By hypothesis and continuity in time, this does not hold, hence $\mu_t^N(f)$ does not converge a.s. to $\mu_t(f)$ as $N \to \infty$ and, in particular, the bound (4.40) does not hold.

□

This result is in disagreement with the convergence results for discrete time resampling algorithms (see e.g. Douc et al., 2005). One possible explanation is that in discrete time the sampling rate λ^Δ is effectively of order N if resampling happens at every time step together with mutation. In continuous time, mutation events occur independently for each clone, so in total at rate N, while resampling only has a rate of order 1. While it is natural that we cannot expect a central limit theorem for resampling algorithms given by (4.33), different versions of resampling could work as, for example, performing resampling only during mutation events. However, for the definition of resampling algorithms itself, the number of particles that get replaced during a single resampling event is typically not bounded uniformly in N and thus condition (3.3c), i.e.

$$\sup_{t\geq 0} J_t^N := \sup_{t\geq 0} |\{i \in 1, \ldots, N \, : \, \xi_t^i \neq \xi_{t-}^i\}| \; \leq \; K \;\; \text{a.s.} \; , \quad K \text{ indep. of } N \; ,$$

is not satisfied. This condition is essential for applying Lemma 3.2.1 in the proof of L^p convergence, however it could be possible to replace it with a weaker condition on the jumps. For instance, one possible weaker assumption is to require that the expectation $\mathbb{E}\big[\sup_{s\leq T} |J_s^N|^p\big]$ is bounded for any $p \geq 1$. This would still give a similar result to Lemma 3.2.1 (see, for instance, Borovskikh and Korolyuk, 2019, Theorem 3.2.4, for discrete-time martingales).

Chapter 5

Application to Rare Events Simulation

Dynamic large deviations of continuous-time jump processes are a common application area of cloning algorithms (Giardinà et al., 2006; Lecomte and Tailleur, 2007). In this Chapter, we present how to apply interacting particle approximations to the study of large deviations for additive path functionals of pure jump processes. In particular, we will present numerical procedures for evaluating the *scaled cumulant generating function* (SCGF), which plays an essential role in the study of large deviations. We conclude the chapter by applying a mean field particle approximation and a cloning algorithm to the inclusion process as an example.

5.1 Large deviation theory

In this section, we introduce the key elements of Large Deviation Theory needed to define and study the large deviations for additive path functionals of pure jump processes. The main result stated here is Theorem 5.1.7, which motivates the importance of the SCGF (Definition 5.1.4) in the study of the large deviations. The main references for this section are Dembo and Zeitouni (2009) and Den Hollander (2008).

Throughout this section, we consider a family of bounded continuous real-valued functions $h_t \in \mathcal{C}_b(E)$, $t \geq 0$, and consider a Markov process $(X_t)_{t\geq 0}$ on a Polish state space E, with initial distribution $\mu_0 \in \mathcal{P}(E)$. The large deviation principle (LDP) characterises the limiting behaviour as $t \to \infty$ of $\mathbb{P}_{\mu_0}(h_t(X_t) \in \cdot)$ in terms of a rate function.

Definition 5.1.1 (Rate function)**.** The function $I : \mathbb{R} \to [0, \infty]$ is called a *rate function* if:

(D1) $I \not\equiv \infty$.

(D2) I is lower semi-continuous.

(D3) I has compact level sets, that is $\{x \in \mathbb{R} \,|\, I(x) \leq c\}$ is compact for all $c \in [0, \infty)$.

Remark. In the large deviation literature, functions $I : \mathbb{R} \to [0, \infty]$ satisfying conditions (D1)-(D3) are sometimes referred to as *good rate functions* (Dembo and Zeitouni, 2009), to distinguish them from functions $I : \mathbb{R} \to [0, \infty]$ that only satisfy conditions (D1)-(D2). Since in this work we always assume that the rate functions have compact level sets as in Definition 5.1.1, there is no confusion with the notation and we simply refer to them as 'rate functions'.

Given a rate function $I : \mathbb{R} \to [0, \infty]$, we define the corresponding set function by

$$I(S) = \inf_{x \in S} I(x), \quad S \subset \mathbb{R}.$$

Definition 5.1.2 (Large deviation principle)**.** The family $\left(\mathbb{P}_{\mu_0}(h_t(X_t) \in \cdot)\right)_{t \geq 0}$ is said to satisfy the *large deviation principle* (LDP) with rate function I if:

(D1') I is a rate function in the sense of Definition 5.1.1;

(D2') $\limsup_{t \to \infty} \frac{1}{t} \log \mathbb{P}_{\mu_0}(h_t(X_t) \in C) \leq -I(C)$, for all $C \subset \mathbb{R}$ closed;

(D3') $\liminf_{t \to \infty} \frac{1}{t} \log \mathbb{P}_{\mu_0}(h_t(X_t) \in O) \geq -I(O)$, for all $O \subset \mathbb{R}$ open.

Lemma 5.1.3. *If $\left(\mathbb{P}_{\mu_0}(h_t(X_t) \in \cdot)\right)_{t \geq 0}$ satisfies the LDP, then the associated rate function I is unique.*

Proof. See Dembo and Zeitouni (2009, Lemma 4.1.4). □

Remark. The LDP implies that

$$\inf_{x \in \mathbb{R}} I(x) = I(\mathbb{R}) = 0 \ ,$$

because I is non-negative by definition and $\mathbb{P}_{\mu_0}(h_t(X_t) \in \mathbb{R}) = 1$.

Moreover, there exists an $x \in \mathbb{R}$ such that $I(x) = 0$, since a lower semi-continuous function attains a minimum on every non-empty compact set and I has compact level sets.

In what follows we show that, under general conditions, the rate function associated to an LDP is the Fenchel-Legendre transform of the scaled cumulant generating function (SCGF). This motivates our interest in approximating the SCGF via interacting particle approximations in order to study the large deviations associated to a pure jump Markov process.

The *logarithmic moment generating function* $\Lambda_t : \mathbb{R} \to (-\infty, \infty]$ is defined to be

$$\Lambda_t(k) := \log \mathbb{E}_{\mu_0}\left[e^{k\,h(X_t)}\right], \quad k \in \mathbb{R}\ .$$

Definition 5.1.4. The *scaled cumulant generating function (SCGF)* is defined by

$$\Lambda(k) := \lim_{t\to\infty} \frac{1}{t}\,\Lambda_t(k \cdot t)\ ,$$

provided the limit exists.

Definition 5.1.5. In this setting, the *Fenchel-Legendre transform* of a function $f : \mathbb{R} \to [-\infty, \infty]$ is defined as

$$f^\star(a) := \sup_{x\in\mathbb{R}} \{a\,x\ -\ f(a)\}\ , \quad a \in \mathbb{R}\ .$$

Lemma 5.1.6 (Duality Lemma)**.** *Let $f : \mathbb{R} \to (-\infty, \infty]$ be a lower semi-continuous, convex function and let $g(x) := f^\star(x)$, for any $x \in \mathbb{R}$. Then also $f(x) = g^\star(x)$.*

Proof. See Dembo and Zeitouni (2009, Lemma 4.5.8).

□

The key application of the duality lemma is in the following result which enables us to identify convex rate functions as the Fenchel-Legendre transform of SCGFs.

Theorem 5.1.7. *Assume that $\mathbb{P}_{\mu_0}(h_t(X_t) \in \cdot)$ satisfies the LDP with rate function $I : \mathbb{R} \to [0, \infty]$ and that*

$$\limsup_{t\to\infty} \frac{1}{t}\,\Lambda_t(k\,t)\ <\ \infty\ , \quad \text{for any } k \in \mathbb{R}\ .$$

Then,

1. *For each $k \in \mathbb{R}$, the limit $\Lambda(k)$ exists, is finite and satisfies*

$$\Lambda(k) = \sup_{x\in\mathbb{R}}\{k\,x\ -\ I(x)\}\ ;$$

2. *If I is convex, then it is the Fenchel-Legendre transform of Λ, namely $I(x) = \Lambda^\star(x)$;*

3. *If I is not convex, then $\Lambda^\star$ is the affine regularisation of I, i.e. $\Lambda^\star \leq I$ and $f \leq \Lambda^\star$, for any convex rate function f such that $f \leq I$.*

Proof. See Dembo and Zeitouni (2009, Theorem 4.5.10). □

Remark. In statistical mechanics, Theorem 5.1.7 corresponds to the thermodynamical equivalence between canonical and microcanonical ensembles (Touchette, 2003, Section III). The SCGF plays indeed an essential role in the investigation of non-equilibrium systems - a role similar to the free energy in equilibrium ones (Touchette, 2009), whereas the rate function I plays a role akin to the specific entropy.

5.2 Large deviations and Feynman-Kac models

For a given process $(X_t)_{t\geq 0}$ with bounded rates $W(x, dy) = \lambda(x)p(x, dy)$ and path space $\Omega = D([0, \infty), E)$ as outlined in Section 2.1, we consider a time-additive observable $A_T : \Omega \to \mathbb{R}$, taken to be a real measurable function of the paths of X_t over the time interval $[0, T]$ of the form

$$A_T(\omega) := \frac{1}{T} \sum_{\substack{t\leq T \\ \omega(t_-)\neq\omega(t)}} g\big(\omega(t_-), \omega(t)\big) + \frac{1}{T}\int_0^T h\big(\omega(t)\big)dt \,. \tag{5.1}$$

Here $g \in \mathcal{C}_b(E^2)$ is such that $g(x, x) = 0$, for any $x \in E$, and $h \in \mathcal{C}_b(E)$, with $\omega \in \Omega$ a realisation of $(X_t : t \geq 0)$. Note that A_T is well defined since the bound on $\lambda(x)$ implies that the process does not explode and the first sum contains almost surely only finitely many non-zero terms for any $T \geq 0$.

Example 5.2.1. The considered class of observables A_T includes many random variables of mathematical and physical interest, such that:

- the occupation time in some set $B \subset E$, which can be obtained by considering $g \equiv 0$ and $h(x) = \mathbb{1}_B(x)$, where $\mathbb{1}_B$ is the characteristic function of B;
- the net flux on a finite state space $E \subset \mathbb{Z}^d$ across a particular bond (i, j), with $||i - j|| = 1$, obtained with $g(x, y) = \mathbb{1}_{\{i\}}(x) \cdot \mathbb{1}_{\{j\}}(y) - \mathbb{1}_{\{j\}}(x) \cdot \mathbb{1}_{\{i\}}(y)$ and $h(x) \equiv 0$;

- the action functional (Lebowitz and Spohn, 1999) on finite state spaces with jump rates $W(i,j)$, obtained by setting $g(x,y) = \log \frac{W(x,y)}{W(y,x)}$ and $h(x) \equiv 0$, which is related to the entropy production in non-equilibrium systems.

More precisely, we are interested in studying the limiting behaviour, as $T \to \infty$, of the family of probability measures $\mathbb{P}_{\mu_0}(A_T \in \cdot\,) = \mathbb{P}_{\mu_0} \circ A_T^{-1}$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, where μ_0 represents the initial distribution of the underlying process. This can be characterised by the *large deviation principle* (LDP). Throughout this chapter, we assume that an LDP with convex rate function I holds. This has been established by Bertini et al. (2015) for a large class of functionals of type (5.1), which includes all our examples of interest.

Adapting the results in Section 5.1 to probability path measures on Ω, we know that in the study of the long-time limit behaviour of $\mathbb{P}_{\mu_0}(A_T \in \cdot\,)$, a key role is played by the scaled cumulant generating function (SCGF) of the observable A_T, i.e.

$$\Lambda(k) = \lim_{T\to\infty} \frac{1}{T} \log \mathbb{E}_{\mu_0}\left[e^{kTA_T}\right] , \tag{5.2}$$

with $k \in \mathbb{R}$. Indeed, by Theorem 5.1.7, if the rate function I is convex and

$$\limsup_{T\to\infty} \frac{1}{T} \log \mathbb{E}_{\mu_0}\left[e^{kTA_T}\right] < \infty ,$$

for every $k \in \mathbb{R}$, then I is fully characterised by the SCGF via Legendre duality, i.e.

$$\Lambda(k) = \sup_{a\in\mathbb{R}}\{k\,a - I(a)\} \quad \text{and} \quad I(a) = \sup_{k\in\mathbb{R}}\{k\,a - \Lambda(k)\}.$$

The SCGF is also the object that can be numerically approximated by cloning algorithms (Giardinà et al., 2006; Lecomte and Tailleur, 2007) and related approaches, and our main aim in this chapter is to illustrate how our results on Feynman-Kac models can be applied here. Thus, we restrict ourselves to settings where $\Lambda(k)$ exists and is finite. In the following we introduce the associated Feynman-Kac models in the notation that is established in this context.

Lemma 5.2.2. *For any $k \in \mathbb{R}$ the family of operators $\big(P_k(t) : t \geq 0\big)$ on $\mathcal{C}_b(E)$ defined by*

$$P_k(t)f(x) := \mathbb{E}_x\big[f\big(X_t\big)\, e^{ktA_t}\big], \tag{5.3}$$

with $f \in \mathcal{C}_b(E)$, is well defined and it is a non-conservative semigroup, the so-called tilted semigroup.

Moreover, the infinitesimal generator associated with $\big(P_k(t) : t \geq 0\big)$, in the

sense of the Hille-Yosida Theorem, can be written in the form

$$\mathcal{L}_k(f)(x) = \int_E W(x,dy)[e^{kg(x,y)}f(y) - f(x)] + kh(x)f(x), \tag{5.4}$$

for $f \in \mathcal{C}_b(E)$ and all $x \in E$, with g and h the bounded continuous functions which characterise A_T via (5.1). In particular, the semigroup $P_k(t)$ satisfies the differential equations

$$\frac{d}{dt}P_k(t)f \;=\; P_k(t)\,\mathcal{L}_k(f) \;=\; \mathcal{L}_k\big(P_k(t)f\big), \tag{5.5}$$

for all $f \in \mathcal{C}_b(E)$ and $t \geq 0$.

Proof. See Chetrite and Touchette (2015), Appendix A.1. □

Observe that, if the SCGF (5.2) is independent of the choice of the initial distribution μ_0, it can be written in terms of the tilted semigroup as

$$\Lambda(k) = \lim_{t\to\infty} \frac{1}{t} \log\big(P_k(t)\,1(x)\big), \tag{5.6}$$

for all $x \in E$. However, the tilted semigroup $P_k(t)$ does not conserve probability and therefore it does not provide a corresponding process to sample from and use standard MCMC methods to estimate the SCGF Λ_k. This can be achieved by interpreting the tilted generator $\mathcal{L}_k$ through Feyman-Kac models in a similar fashion to $\mathcal{L}^{\mathcal{V}}$ in Lemma 2.2.1, so that we can apply our results from Section 2.2 and construct the corresponding McKean models.

Lemma 5.2.3. *The infinitesimal generator $\mathcal{L}_k$ (5.4) can be written as*

$$\mathcal{L}_k(f)(x) \;=\; \widehat{\mathcal{L}}_k(f)(x) \;+\; \mathcal{V}_k(x)\cdot f(x)\;, \tag{5.7}$$

for all $f \in \mathcal{C}_b(E)$ and $x \in E$. Here

$$\widehat{\mathcal{L}}_k(f)(x) := \int_E W(x,dy)e^{kg(x,y)}\big(f(y) - f(x)\big) \tag{5.8}$$

is the generator of a pure jump process with modified rates $W(x,dy)\,e^{kg(x,y)}$, and

$$\mathcal{V}_k(x) := \widehat{\lambda}_k(x) - \lambda(x) + kh(x) \;\in\; \mathcal{C}_b(E), \tag{5.9}$$

is a diagonal potential term where $\widehat{\lambda}_k(x) := \int_E W(x,dy)e^{kg(x,y)}$ is the escape rate of $\widehat{\mathcal{L}}_k$.

Proof. Follows directly from the definition of $\mathcal{L}_k$ in (5.4), by simply observing that

$$\begin{aligned}\mathcal{L}_k(f)(x) &= \int_E W(x,dy)\left(e^{kg(x,y)}f(y) - f(x)\right) + kh(x)f(x) \\ &= \int_E W(x,dy)e^{kg(x,y)}\cdot\left(f(y)-f(x)\right) + f(x)\cdot\left(\widehat{\lambda}(x) - \lambda(x) + kh(x)\right),\end{aligned}$$

since $\widehat{\lambda}_k(x) = \int_E W(x,dy)e^{kg(x,y)}$ and $\lambda_k(x) = \int_E W(x,dy)$, by definition.

□

Lemma 5.2.3 states in particular that the tilted semigroup $P_k(t)$ (5.3) is a Feynman-Kac semigroup. Moreover, denoting by $\widehat{X}_k(t)$ the Markov process with infinitesimal generator $\widehat{\mathcal{L}}_k$, the tilted semigroup $P_k(t)$ can be also expressed in terms of the process $\widehat{X}_k(t)$ and the potential $\mathcal{V}_k$ via the standard formula

$$P_k(t)f(x) \;=\; \mathbb{E}_x\Big[f(\widehat{X}_k(t))\cdot\exp\Big(\int_0^t \mathcal{V}_k(\widehat{X}_k(s))\,ds\Big)\Big]\,.$$

As shown in Section 2.2, in order to control the asymptotic behaviour of the tilted semigroup $P_k(t)$, we require that Assumption 2.2.2 on asymptotic stability is satisfied. By (5.6), the SCGF $\Lambda(k)$ is the spectral radius of the generator $\mathcal{L}_k$ (see also (5.11) below). With Assumption 2.2.2 on asymptotic stability, $\Lambda(k)$ is also the principal eigenvalue of $\mathcal{L}_k$ and there exist a probability measure $\mu_\infty = \mu_{\infty,k} \in \mathcal{P}(E)$ and constants $\alpha > 0$ and $\rho \in (0,1)$ such that

$$\left\|e^{-t\Lambda(k)}P_k(t)f(\cdot) - \mu_\infty(f)\right\| \le \|f\|\cdot\alpha\rho^t, \tag{5.10}$$

for every $t \ge 0$ and $f \in \mathcal{C}_b(E)$. Note that this implies the independence of the SCGF from the initial distribution, μ_0, and thus (5.6) holds for every initial state $x \in E$. Note that (5.10) implies in particular that $\mu_0 e^{-t\Lambda(k)}P_k(t)$ converges weakly to μ_∞ for all initial distributions μ_0, and that μ_∞ is the unique invariant probability measure for the modified semigroup $t \mapsto e^{-t\Lambda(k)}P_k(t)$. Therefore we have from the generator $\mathcal{L}_k - \Lambda(k)$ of this semigroup that

$$\mu_\infty\big(\mathcal{L}_k(f)\big) = \Lambda(k)\,\mu_\infty(f) \quad \text{for all } f \in \mathcal{C}_b(E)\,. \tag{5.11}$$

To avoid notation overload, we omit writing explicitly the dependence of certain quantities on the fixed parameter k in the rest of this section.

In analogy with the notation (2.4) for generic pure jump processes, in the

following we also use the notation with a probability kernel

$$W(x,dy)e^{kg(x,y)} = \widehat{\lambda}_k(x)\,\widehat{p}_k(x,dy)\ . \tag{5.12}$$

Observe that

$$\mathcal{L}_k(1)(x) = \mathcal{V}_k(x), \tag{5.13}$$

thus, we get with (5.11) another representation of the SCGF,

$$\Lambda(k) = \mu_\infty(\mathcal{V}_k)\ . \tag{5.14}$$

Recall the unnormalised and normalised versions of the Feynman-Kac measures defined in (2.9) and (2.11) for a given initial distribution $\mu_0 \in \mathcal{P}(E)$,

$$\nu_t(f) = \mu_0\big(P_k(t)f\big) \quad \text{and} \quad \mu_t(f) = \nu_t(f)/\nu_t(1)\ , \quad f \in \mathcal{C}_b(E)\ ,$$

and that asymptotic stability (5.10) implies that $\mu_t \to \mu_\infty$ weakly as $t \to \infty$ (by Lemma 2.2.3). This suggests the following finite-time approximations for $\Lambda(k)$.

Proposition 5.2.4. *For any $k \in \mathbb{R}$ and every $t \geq 0$, we have that*

$$\log \mathbb{E}_{\mu_0}\big[e^{ktA_t}\big] \ = \ \int_0^t \mu_s(\mathcal{V}_k)\,ds,$$

where $\mathcal{V}_k$ is defined in (5.9). In particular, if asymptotic stability (5.10) is satisfied,

$$\frac{1}{T}\int_0^T \mu_s(\mathcal{V}_k)\,ds \ \to \ \Lambda(k) \quad \text{as } T \to \infty\ .$$

Proof. Recalling the evolution equation (2.10) of ν_t, we have

$$\frac{d}{dt}\log \nu_t(1) \ = \ \frac{1}{\nu_t(1)} \cdot \frac{d}{dt}\nu_t(1) \ = \ \frac{\nu_t\big(\mathcal{L}_k(1)\big)}{\nu_t(1)} \ = \ \mu_t\big(\mathcal{L}_k(1)\big).$$

And, thus,

$$\nu_t(1) \ = \ \exp\bigg(\int_0^t \mu_s\big(\mathcal{L}_k(1)\big)\,ds\bigg),$$

since $\nu_0(1) = 1$. We can conclude by observing that $\mathcal{L}_k(1)(x) = \mathcal{V}_k(x)$ and

$$\nu_t(1) = \mathbb{E}_{\mu_0}\big[e^{ktA_t}\big], \tag{5.15}$$

using that the SCGF is well defined under asymptotic stability (5.10).

□

For any $t < T$, we define

$$\Lambda^{t,T}(k) := \frac{1}{T-t} \int_t^T \mu_s(\mathcal{V}_k) ds \tag{5.16}$$

as a finite-time approximation for $\Lambda(k)$.

Lemma 5.2.5. *Under asymptotic stability (5.10), there exist constants $\alpha' > 0$ and $0 < \rho < 1$ such that*

$$\left|\Lambda^{aT,T}(k) - \Lambda(k)\right| \leq \|\mathcal{V}_k\| \cdot \frac{\alpha' \, \rho^{aT}}{(1-a)T} \ ,$$

for every $a \in [0, 1)$.

Proof. By (2.12), we have

$$\begin{aligned}
\left| \frac{1}{(1-a)T} \int_{aT}^T \mu_t(\mathcal{V}_k) dt \ - \ \mu_\infty(\mathcal{V}_k) \right| \ &\leq \ \frac{1}{(1-a)T} \int_{aT}^T \left|\mu_t(\mathcal{V}_k) - \mu_\infty(\mathcal{V}_k)\right| dt \\
&\leq \frac{1}{(1-a)T} \int_{aT}^T \|\mathcal{V}_k\| \cdot \tilde{\alpha} \, \rho^t dt \\
&= \frac{\tilde{\alpha} \, \|\mathcal{V}_k\|}{(1-a)T} \cdot \frac{\rho^T - \rho^{aT}}{\log \rho} \\
&\leq \|\mathcal{V}_k\| \cdot \frac{\alpha' \, \rho^{aT}}{(1-a)T} \ ,
\end{aligned}$$

where $\alpha' := \tilde{\alpha}/(-\log \rho) > 0$, using the basic fact $0 \leq \rho^{aT} - \rho^T \leq \rho^{aT}$. In particular, $\lim_{T \to \infty} \Lambda^{aT,T}(k) = \mu_\infty(\mathcal{V}_k) = \Lambda(k)$, by (5.14).

□

Note that for $a = 0$ the above result only implies a convergence rate of order $1/T$, since errors from the arbitrary initial condition have to be averaged out over time. In contrast, for a fixed $a > 0$ which corresponds to the usual idea of burn-in in conventional Markov chain Monte Carlo approximations where the initial iterations of the simulation are discarded (see, for example, Gelman and Shirley, 2011), we get a much better exponential rate of convergence dominated by the asymptotic stability parameter $\rho \in (0, 1)$.

We conclude the section with a simple example to illustrate the construction of the overall transition rates $\widehat{W}$ and $\widetilde{W}_\mu$, with $\mu \in \mathcal{P}(E)$, associated respectively to the mutation and selection dynamics of the McKean model.

Example 5.2.6. Let $S_M = \mathbb{Z}^d \cap [-M, M]^d$, with $M \in \mathbb{N}$, with periodic boundary conditions $M+1 = -M$ and $-M-1 = M$. We consider the continuous-time random

walk defined by overall transition rates $W(x,\, y) = \frac{\lambda}{2d}$ with $\lambda > 0$, if $||x - y|| = 1$, and $W(x,\, y) = 0$ otherwise. We are interested in counting the number of times the random walker crosses a particular bond $(i,\, j)$ with $||j - i|| = 1$. This is obtained by considering an observable A_T in the form (5.1) with

$$g(x, y) := \mathbb{1}_{\{i\}}(x) \cdot \mathbb{1}_{\{j\}}(y) \,+\, \mathbb{1}_{\{j\}}(x) \cdot \mathbb{1}_{\{i\}}(y) \;,$$

and $h \equiv 0$. This observable is also called *activity* and note the difference in sign in its definition compared to the net flux across a bond, illustrated in Example 5.2.1. Given $k \in \mathbb{R}$, by Lemma 5.2.3, the corresponding tilted generator $\mathcal{L}_k$ can be written in the form $\mathcal{L}_k = \widehat{\mathcal{L}}_k + \mathcal{V}_k$, where the mutation generator $\widehat{\mathcal{L}}_k$ is defined by overall transition rates

$$\widehat{W}_k(x, y) \,=\, W(x, y) e^{kg(x,y)} \,=\, \begin{cases} \frac{\lambda}{2d}\, e^k & (x,\, y) \,=\, (i,\, j) \text{ or } (j,\, i) \\ \frac{\lambda}{2d} & \|x - y\| = 1 \text{ and } (x,\, y) \,\neq\, (i,\, j),\, (j,\, i) \\ 0 & \text{otherwise.} \end{cases}$$

Hence, the escape rate for the mutation events is

$$\widehat{\lambda}_k(x) \,=\, \sum_{||y-x||=1} \frac{\lambda}{2d}\, e^{kg(x,y)} \,=\, \begin{cases} \frac{\lambda}{2d}\, e^k \,+\, \frac{\lambda(2d-1)}{2d} & x \in \{i,\, j\} \\ \lambda & \text{otherwise,} \end{cases}$$

and the potential $\mathcal{V}_k$ is

$$\mathcal{V}_k(x) \,=\, \widehat{\lambda}_k(x) \,-\, \lambda(x) \,=\, \begin{cases} \frac{\lambda}{2d}\, (e^k - 1) & x \in \{i,\, j\} \\ 0 & \text{otherwise.} \end{cases}$$

As discussed in Section 2.3, there are different ways to construct the selection part of the corresponding McKean model. One possible construction is given for instance by (2.21b); in this case, if $k > 0$, the selection events are characterised by the rates

$$\widetilde{W}(x,\, y) \,=\, \begin{cases} \frac{\lambda}{2d}(e^k - 1) & x \notin \{i,\, j\} \text{ and } y \in \{i, j\} \\ 0 & \text{otherwise,} \end{cases}$$

for any $x, y \in S_M$; whereas, if $k < 0$, the selection events are characterised by the

rates

$$\widetilde{W}(x,\,y)\;=\;\begin{cases}\frac{\lambda}{2d}(1-e^{k}) & x\in\{i,\,j\}\text{ and }y\notin\{i,j\}\\ 0 & \text{otherwise,}\end{cases}$$

for any $x, y \in S_M$. In the trivial case $k = 0$, $\widetilde{W}(x, y) \equiv 0$ so no selection event occurs and $\widehat{W}_0(x, y) = W(x, y)$, thus the mutation dynamics corresponds to the original process.

5.3 Estimation of the SCGF

In this section we establish the convergence of estimators of the SCGF, $\Lambda(k)$ (5.2), provided by interacting particle approximations. Approximating μ_t by the empirical distribution μ_t^N (3.2) associated to an interacting particle system, we can estimate $\Lambda^{t,T}(k)$ with

$$\Lambda^{t,T,N}(k) := \frac{1}{T-t}\int_t^T \mu_s^N(\mathcal{V}_k)\,ds\ . \tag{5.17}$$

Note that, choosing $f \equiv 1$ in Proposition 3.1.4 about unbiased estimators and recalling (5.15), we see that $\exp\big(t\cdot\Lambda^{0,t,N}(k)\big)$ is an unbiased estimator of $\exp\big(t\cdot\Lambda^{0,t}(k)\big)$. Recall that particle approximations are characterised by a sequence of generators $(\overline{L}^N)_{N\in\mathbb{N}}$ on $\mathcal{C}_b(E^N)$, based on McKean generators (2.16)

$$\overline{\mathcal{L}}_{\mu,k} := \widehat{\mathcal{L}}_k + \widetilde{\mathcal{L}}_{\mu,k} \quad \text{for all } \mu\in\mathcal{P}(E)\ ,$$

where $\widetilde{\mathcal{L}}_{\mu,k}$ describes the selection dynamics of the McKean model as in Lemma 2.3.1, with examples in (2.21a) or (2.21b). Due to tilted dynamics explained in Lemma 5.2.3 we have an additional dependence on the parameter k.

Proposition 5.3.1. *Let $(\overline{L}_k^N)_{N\in\mathbb{N}}$ be a sequence of generators satisfying Assumption 3.1.1 with McKean generators $\overline{\mathcal{L}}_{\mu,k}$. With asymptotic stability (5.10), for every $p \geq 2$ and $a \in [0, 1)$ there exist constants c_p, c', $\alpha' > 0$ and $0 < \rho < 1$ independent of N and T such that*

$$\mathbb{E}\Big[\,\big|\Lambda^{aT,T,N}(k) - \Lambda(k)\big|^p\Big]^{1/p} \leq \frac{c_p}{N^{1/2}} + \frac{\alpha'\cdot\rho^{aT}}{(1-a)T}\ , \tag{5.18}$$

and

$$\Big|\mathbb{E}\Big[\Lambda^{aT,T,N}(k)\Big] - \Lambda(k)\Big| \leq \frac{c'}{N} + \frac{\alpha'\cdot\rho^{aT}}{(1-a)T}\ , \tag{5.19}$$

for any $N \in \mathbb{N}$ large enough and $T > 0$.

Proof. First, note that

$$\mathbb{E}\big[\,\big|\Lambda^{aT,T,N}(k) - \Lambda(k)\big|^p\big]^{1/p} \leq \mathbb{E}\big[\,\big|\Lambda^{aT,T,N}(k) - \Lambda^{aT,T}(k)\big|^p\big]^{1/p} + \big|\Lambda^{aT,T}(k) - \Lambda(k)\big| \,.$$

The bound for the second term is given in Lemma 5.2.5, whereas we can bound the first term by observing that

$$\mathbb{E}\Big[\,\big|\Lambda^{aT,T,N}(k) - \Lambda^{aT,T}(k)\big|^p\Big]^{1/p} \leq \frac{1}{(1-a)T}\int_{aT}^{T} \mathbb{E}\big[\,\big|\mu_t^N(\mathcal{V}_k) - \mu_t(\mathcal{V}_k)\big|^p\big]^{1/p}\,dt \,,$$

and applying Theorem 3.1.2. The second claim can be established similarly.

□

Proposition 5.3.1 provides the L^p and bias estimates of the approximation error with order of convergence respectively given by $1/\sqrt{N}$ and $1/N$. The necessarily finite simulation time T leads to an additional error of order ρ^{aT}/T, with $\rho \in (0,1)$, which is controlled by asymptotic stability properties of the process as summarised in Lemma 5.2.5. Ideally, during simulations we want to choose the final time $T = T(N)$ with respect to the population size N so to balance both terms in (5.18), resp. (5.19). Indeed, choosing $T(N) \gg N$ would be computationally more expensive than $T(N) \approx N$, but would give the same order of convergence. Proposition 5.3.1 also implies that $\Lambda^{aT,T,N}(k)$ converges almost surely to $\Lambda^{aT,T}(k)$ as $N \to \infty$. Similarly, we can adapt Theorem 3.1.3 to obtain a Central Limit Theorem for the estimator $\Lambda^{t,T,N}(k)$ of $\Lambda^{t,T}(k)$.

Proposition 5.3.2. *[CLT] Let $(\overline{L}_k^N)_{N\in\mathbb{N}}$ be a sequence of generators satisfying Assumption 3.1.1 with McKean generators $\overline{\mathcal{L}}_{\mu,k}$. With asymptotic stability (5.10), for any fixed $0 \leq t \leq T$, the sequence*

$$U_{t,T}^N := \sqrt{N}(\Lambda^{t,T,N}(k) - \Lambda^{t,T}(k)) \,,$$

converges in law as $N \to \infty$ to a centered Gaussian random variable $U_{t,T}$ whose variance is given by

$$\mathbb{E}\big[U_{t,T}^2\big] \;=\; \frac{1}{T-t}\,\mathbb{E}\Big[\Big(\int_t^T V_s(\mathcal{V}_k)\,ds\Big)^2\Big] \,,$$

with V_s given in Theorem 3.1.3.

Proof. By construction (5.17) of the estimator $\Lambda_k^{t,T,N}$,

$$U_{t,T}^N = \frac{1}{T-t}\int_t^T \sqrt{N}\left(\mu_s^N(\mathcal{V}_k) - \mu_s(\mathcal{V}_k)\right) ds \ ,$$

then the statement follows straightforwardly from the Central Limit Theorem for the empirical measures μ_s^N (Theorem 3.1.3).

□

Proposition 5.3.2 also provides an explicit formulation of the asymptotic variance for an interacting particle approximation of the SCGF, however it is not numerically tractable and does not provide any useful insight in the comparison of the asymptotic variance for two different particle approximation methods. In order to obtain better results, one would need to evaluate covariances for $V_s(\mathcal{V}_k)$, i.e. $\mathbb{E}[V_s(\mathcal{V}_k) \cdot V_u(\mathcal{V}_k)]$ for any $t \leq s, u \leq T$, and this requires a stronger result than the CLT for empirical measures μ_s^N (Theorem 3.1.3).

5.4 The cloning factor

Most results in the physics literature do not use the estimator $\Lambda^{aT,T,N}(k)$ (5.17) based on the ergodic average of the mean fitness of the clone ensemble, but an estimator based on a so-called 'cloning factor' (see, e.g., Giardinà et al., 2006, 2011; Pérez-Espigares and Hurtado, 2019). This is essentially a continuous-time jump process $(C_t^N : t \geq 0)$ on $(0, \infty)$ with $C_0^N = 1$, where at each cloning event of size $n \in \mathbb{N}_0 \cup \{-1\}$ at a given time τ, the value is updated as

$$C_t^N = C_{t-}^N \left(1 + \frac{n}{N}\right) ,$$

where $n = -1$ occurs when there is a 'killing' event. In this section, we present convergence results for the estimator based on the cloning factor and show that the L^p-error for this estimator has the same rate of convergence $1/\sqrt{N}$ as $\Lambda^{aT,T,N}(k)$ (5.17).

In line with the physics literature (Lecomte and Tailleur, 2007), the dynamics of the cloning factor C_t^N is defined jointly with the cloning algorithm given by (4.31) via an extension of the cloning generator $\overline{L}_{c,k}^N$, based on the McKean model (2.21a) with parameter $c \in \mathbb{R}$, with exit rate $\lambda(x)$ and probability kernel $p(x, dy)$ replaced

by $\widehat{\lambda}_k(x)$ and $\widehat{p}_k(x, dy)$ respectively, i.e.

$$\overline{L}^N_{c,k}(F)(\underline{x}) \;=\; \sum_{i=1}^N \Big(\widehat{\lambda}_k(x_i) \int_E \widehat{p}_k(x_i,\, dy) \sum_{A\in\mathcal{N}} \pi_{\underline{x}}(x_i,\, A) \cdot \big(F(\underline{x}^{A,x_i;\,i,y}) - F(\underline{x})\big) \\ + \sum_{j=1}^N \frac{\big(\mathcal{V}_k(x_i) - c\big)^-}{N} \big(F(\underline{x}^{i,x_j}) - F(\underline{x})\big) \Big) \;.$$

Thus, the joint dynamics on the state space $E^N \times (0,\infty)$ is given by the generator

$$\overline{L}^{(N,\star)}_{c,k}(F^\star)(\underline{x},\varsigma) := \\ \sum_{i=1}^N \Big(\widehat{\lambda}_k(x_i) \int_E \widehat{p}_k(x_i,\, dy) \sum_{A\in\mathcal{N}} \pi_{\underline{x}}(x_i,\, A) \big(F^\star(\underline{x}^{A,x_i;\,i,y}, \varsigma_{|A|}) - F^\star(\underline{x},\varsigma)\big) \\ + \sum_{j=1}^N \frac{\big(\mathcal{V}_k(x_i) - c\big)^-}{N} \big(F^\star(\underline{x}^{i,x_j}, \varsigma_{-1}) - F^\star(\underline{x},\varsigma)\big) \Big) \;, \tag{5.20}$$

where the test function $F^\star : E^N \times (0,\infty) \to \mathbb{R}$ now has a second counting coordinate, and we denote $\varsigma_n := \varsigma \cdot \big(1+\frac{n}{N}\big)$, with $n \in \mathbb{N}_0 \cup \{-1\}$.

To simplify the presentation, we further assume that the probability of choosing a set A depends only on its size $|A|$ and not on its elements, i.e. for any $x \in E$

$$\pi_{\underline{x}}(x, A) = \widetilde{\pi}(x,\, |A|) \Big/ \binom{N}{|A|} \quad \text{with} \quad \widetilde{\pi}(x,\, 0), \ldots, \widetilde{\pi}(x,\, N) \text{ s.t. } \sum_{n=0}^N \widetilde{\pi}(x,\, n) = 1 \\ \text{and} \quad \widetilde{\pi}(x,\, n) = \delta_{n,0} \text{ if } \mathcal{V}(x) \le c \;. \tag{5.21}$$

Using this construction, we have

$$\psi_{\underline{x}}(x_i,\, x_j) \;=\; \sum_{n=1}^N \frac{n}{N} \widetilde{\pi}(x_i,\, n) \;=:\; \frac{M(x_i)}{N} \;.$$

Thus, conditions (4.14)-(4.17) are satisfied by assuming

$$M(x) \;=\; \frac{\big(\mathcal{V}_k(x) - c\big)^+}{\widehat{\lambda}_k(x)} \;. \tag{5.22}$$

We introduce the coordinate projection $G(\underline{x},\, \varsigma) := \varsigma$ in order to observe only the cloning factor, $G(\zeta^N_t,\, C^N_t) = C^N_t$. Note that $E^N \times (0,\infty)$ is not compact, and G is an unbounded test function. However, since the range of the clone size distribution is uniformly bounded (condition 3.3c), $t \mapsto \log C^N_t$ is a birth-death process on $[0,\infty)$

with bounded jump length, and the generator (5.20) and associated semigroup are therefore well defined for the test function G (by Lemma 2.1.10) and all $t \geq 0$.

The following result provides an unbiased estimator for the unnormalised quantity $\nu_t(1)$ based on the cloning factor.

Proposition 5.4.1. *Let $\overline{L}^{(N,\star)}_{c,k}$ be the extension (5.20) of the cloning generator $\overline{L}^{N}_{c,k}$ (4.13). Then, the quantity $e^{tc}C^N_t$ is an unbiased estimator for $\nu_t(1)$ (2.9), i.e.*

$$\mathbb{E}\big[e^{tc}C^N_t\big] \;=\; \mathbb{E}\big[\nu^N_t(1)\big] \;=\; \nu_t(1),$$

for every $t \geq 0$ and $N \geq 1$, and all choices of the parameter $c \in \mathbb{R}$ (cf. (2.21a)).

Proof. First observe that, following (4.16),

$$\begin{aligned}\overline{L}^{(N,\star)}_{c,k}(G)(\underline{x},\varsigma) &= \sum_{i=1}^{N}\sum_{n=0}^{N} \widehat{\lambda}_k(x_i)\,\widetilde{\pi}(x_i,\,n)\cdot(\varsigma\, n/N) \;-\; \sum_{i=1}^{N}\frac{\varsigma}{N}\big(\mathcal{V}_k(x_i)-c\big)^{-} \\ &= \frac{\varsigma}{N}\sum_{i=1}^{N}\big(\mathcal{V}_k(x_i)-c\big)\;, \end{aligned} \tag{5.23}$$

using the mean $M(x_i)$ of the distribution $\widetilde{\pi}(x_i,\,n)$ as given in (5.22). Therefore,

$$\overline{L}^{(N,\star)}_{c,k}(G)(\underline{x},\varsigma) \;=\; \varsigma\, m(\underline{x})\big(\mathcal{V}_k - c\big)\;,$$

and analogously to (3.10), the expected time evolution of C^N_t is then given by

$$\frac{d}{dt}\mathbb{E}[C^N_t] \;=\; \mathbb{E}[C^N_t\cdot\mu^N_t(\mathcal{V}_k-c)].$$

This is also the evolution of $\nu^N_t(e^{-tc}) = e^{-tc}\nu^N_t(1)$, since

$$\begin{aligned}\frac{d}{dt}\mathbb{E}[\nu^N_t(e^{-tc})] \;&=\; \mathbb{E}[\mu^N_t(\mathcal{V}_k)\cdot e^{-tc}\nu^N_t(1) - c\; e^{-tc}\nu^N_t(1)] \\ &=\; \mathbb{E}[\nu^N_t(e^{-tc})\cdot\mu^N_t(\mathcal{V}_k-c)].\end{aligned}$$

With initial conditions $C^N_k(0) = 1 = \nu^N_t(1)$, the statement follows by a Gronwall argument analogous to (3.9) and by Proposition 3.1.4.

□

Proposition 5.4.1 leads to an alternative estimator for $\Lambda^{t,T}(k)$ (5.16) given by

$$\overline{\Lambda}^{t,T,N}(k) := \frac{1}{T-t}\big(\log C^N_T - \log C^N_t\big) \;+\; c. \tag{5.24}$$

Note that this is not itself unbiased as a consequence of the nonlinear transformation involving the logarithm.

In order to study the convergence of the new estimator to the SCGF, it is convenient to use the martingale characterisation of the process, which is given by the following result.

Proposition 5.4.2. *Let $\overline{L}_{c,k}^{(N,\star)}$ be the extension (5.20) of the cloning generator $\overline{L}_{c,k}^{N}$. Then, the process*

$$\mathcal{M}_t^\star := \log C_t^N - \int_0^t \overline{L}_{c,k}^{(N,\star)}(H)\big(\zeta_s^N,\, C_s^N\big)\, ds,$$

with $H(\underline{x}, \varsigma) = \log \varsigma$, is a martingale satisfying

$$\mathcal{M}_t^\star = \log C_t^N - \int_0^t \big(\mu_s^N(\mathcal{V}_k) - c\big)\, ds \,+\, t \cdot O\Big(\frac{1}{N}\Big),$$

and with predictable quadratic variation

$$\langle \mathcal{M}_\cdot^\star \rangle_t \,=\, \frac{1}{N} \int_0^t \mu_s^N\big(\widehat{\lambda}_k\, Q \,+\, (\mathcal{V}_k - c)^-\big)\, ds \,+\, t \cdot O\Big(\frac{1}{N^2}\Big)\,,$$

where $Q(x) := \sum n^2\, \widetilde{\pi}(x,\, n)$ is the second moment of the distribution $\widetilde{\pi}(x,\, n)$.

Remark. Note that, in case there is at most one clone per transition event, i.e. if $Q(x_i) = M(x_i) = (\mathcal{V}_k(x_i) - c)^+ / \widehat{\lambda}_k(x_i)$, then

$$\langle \mathcal{M}_\cdot^\star \rangle_t \,=\, \frac{1}{N} \int_0^t \big(\mu_s^N(\mathcal{V}_k) - c\big)\, ds \,+\, t \cdot O\Big(\frac{1}{N^2}\Big)\,.$$

Proof. Observe that we can rewrite (5.20) as

$$\begin{aligned}
\overline{L}_{c,k}^{(N,\star)}(H)(\underline{x}, \varsigma) &= \sum_{i=1}^{N} \left(\sum_{n=0}^{N} \widehat{\lambda}_k(x_i)\, \widetilde{\pi}(x_i,\, n) \log(1{+}n/N) \,+\, \big(\mathcal{V}_k(x_i){-}c\big)^- \log(1{-}1/N) \right) \\
&= m(\underline{x})\big(\mathcal{V}_k\big) - c \,+\, O\Big(\frac{1}{N}\Big)\,,
\end{aligned}$$

using the expansion $\log(1+x) = x + O(x^2)$ as $x \to 0$. Similarly,

$$\begin{aligned}
&\Gamma_{\overline{L}_{c,k}^{(N,\star)}}(H,H)(\underline{x},\varsigma) \\
&\quad = \sum_{i=1}^{N}\Bigg(\sum_{n=0}^{N}\widehat{\lambda}_k(x_i)\,\widetilde{\pi}(x_i,\,n)\,\big(\log(1{+}n/N)\big)^2 \;+\; \big(\mathcal{V}_k(x_i){-}c\big)^{-}\,\big(\log(1{-}1/N)\big)^2\Bigg) \\
&\quad = \frac{1}{N}\,m(\underline{x})\big(\,\widehat{\lambda}_k Q \;+\; (\mathcal{V}_k - c)^{-}\big) \;+\; O\Big(\frac{1}{N^2}\Big).
\end{aligned}$$

The statement corresponds to the martingale problem associated to $\overline{L}_{c,k}^{(N,\star)}(H)$. □

By Proposition 5.4.2 and recalling the definition of the SCGF estimators $\Lambda^{t,T,N}(k)$ (5.17) and $\overline{\Lambda}^{t,T,N}(k)$ (5.24) we immediately get

$$\Lambda^{t,T,N}(k) \;=\; \overline{\Lambda}^{t,T,N}(k) \;-\; \frac{\mathcal{M}_T^\star - \mathcal{M}_t^\star}{T-t} \;+\; O\Big(\frac{1}{N}\Big)\,.$$

In what follows, we discuss the convergence of the estimator $\overline{\Lambda}^{aT,T,N}(k)$ to the SCGF $\Lambda(k)$, which is based on the cloning factor.

Theorem 5.4.3. *Let $\overline{\Lambda}^{t,T,N}(k)$ be the cloning factor estimator for $\Lambda^{t,T}(k)$ as given in (5.24), and let $\Lambda^{t,T,N}(k)$ be the estimator for $\Lambda^{t,T}(k)$ as given in (5.17). Then, for every $p \geq 2$ and $a \in [0,1)$, there exists a constant $C_p^\star > 0$ such that, for all N large enough and $T \geq 1$,*

$$\mathbb{E}\Big[\Big|\overline{\Lambda}^{aT,T,N}(k) - \Lambda^{aT,T,N}(k)\Big|^p\Big]^{1/p} \;\leq\; \frac{C_p^\star}{N^{1+1/p}\cdot\sqrt{T}}. \tag{5.25}$$

If in addition Assumption (5.10) on asymptotic stability holds, there exist constants $\gamma_p^\star$, $c_p^\star, \alpha' > 0$ and $0 < \rho < 1$ (dependent on a, p, $\widehat{\lambda}_k$, Q and $\mathcal{V}_k$) such that

$$\mathbb{E}\Big[\Big|\overline{\Lambda}^{aT,T,N}(k) - \Lambda(k)\Big|^p\Big]^{1/p} \;\leq\; \frac{\gamma_p^\star}{N^{1+1/p}\cdot\sqrt{T}} \;+\; \frac{c_p^\star}{\sqrt{N}} \;+\; \frac{\alpha'\rho^{aT}}{(1-a)T},$$

for every $T \geq 1$.

Proof. Thanks to Jensen's inequality, it is enough to prove the inequality for all

$p = 2^q$, $q \in \mathbb{N}$. First, we can write

$$\begin{aligned}\mathbb{E}\Big[\Big|\overline{\Lambda}^{aT,T,N}(k) - \Lambda^{aT,T,N}(k)\Big|^{2^q}\Big] &= \frac{1}{\big(N\cdot(1-a)T\big)^{2^q}}\cdot\mathbb{E}\Big[\big|\mathcal{M}^\star_T - \mathcal{M}^\star_{aT}\big|^{2^q}\Big] \\ &\leq \frac{1}{\big(N\cdot(1-a)T\big)^{2^q}}\cdot\mathbb{E}\Big[\big|\mathcal{M}^\star_T\big|^{2^q}\Big].\end{aligned}$$

Observe that $\sup_{t\leq T}\big|\mathcal{M}^\star_t\big| < \infty$, so the assumptions of Lemma 3.2.1 are satisfied. Thus, using Lemma 3.2.1, we obtain

$$\begin{aligned}\frac{1}{N^{2^q}\cdot T^{2^q}}\cdot\mathbb{E}\Big[\big|\mathcal{M}^\star_T\big|^{2^q}\Big] &\leq \frac{C_q}{N^{2^q}\cdot T^{2^q}}\sum_{k=0}^{q-1}\mathbb{E}\big[(\langle\mathcal{M}^\star_\cdot\rangle_T)^{2^k}\big] \\ &\leq \frac{\widetilde{C}_q}{N^{2^q}}\sum_{k=0}^{q-1}\frac{1}{T^{2^q-2^k}}\left(\frac{1}{N^{2^k}} + O\Big(\frac{1}{N^{2^k+1}}\Big)\right) \\ &\leq \frac{C^\star_q}{N^{2^q+1}\cdot T^{2^{q-1}}}\ ,\end{aligned}$$

for every $T \geq 1$. The second part of the Theorem follows directly by Proposition 5.3.1.

□

Therefore, the L^p-error for estimator $\overline{\Lambda}^{aT,T,N}(k)$ has the same rate of convergence $1/\sqrt{N}$ as $\Lambda^{aT,T,N}(k)$. Analogous results hold for the bias estimates, which have order of convergence $1/N$ as for the estimator $\Lambda^{aT,T,N}(k)$ (Proposition 5.3.1), since with (5.25) the difference of both estimators is only of order $1/N^{1+1/p}$.

5.5 Current large deviations for lattice gases

In this Section, we illustrate how to apply interacting particle approximations for studying the current large deviations for lattice gases and we apply these algorithms to the inclusion process as an example.

We consider one-dimensional stochastic lattice gases with periodic boundary conditions on the discrete torus $\mathbb{T}_L$ with L sites and a fixed number of particles $M < L$, so that the total density is $\rho_0 = M/L$. Within our general framework, they are simply Markov chains on the finite state space $E = S^{\mathbb{T}_L}$, with $S \subseteq \mathbb{N}_0$, of all particle configurations, which have been of recent research interest in the context of current fluctuations. We denote the configurations by $\eta = (\eta_x \,|\, x \in \mathbb{T}_L)$ where $\eta_x \in \mathbb{N}_0$ is interpreted as the occupation number at site x, and the process is denoted as $\big(\eta(t)\big)_{t\geq 0}$. In order to use standard notation for lattice gases, in this

section $x, y \in \mathbb{T}_L$ denote the sites on the discrete torus, whereas in the previous sections they denoted states in E. Monomers jump to nearest neighbour sites with rates $u(\eta_x, \eta_y) \geq 0$ for $y = x \pm 1$ depending on the occupation numbers of departure and target site, multiplied with a spatial bias $p = 1 - q \in [0, 1]$. The generator is of the form

$$\mathcal{L}f(\eta) = \sum_{x \in \mathbb{T}_L} \Big[p\, u(\eta_x, \eta_{x+1}) \big(f(\sigma_{x,x+1}\eta) - f(\eta) \big) + q\, u(\eta_x, \eta_{x-1}) \big(f(\sigma_{x,x-1}\eta) - f(\eta) \big) \Big] , \tag{5.26}$$

where $\sigma_{x,y}\eta$ results from the configuration η after moving one particle from x to y. The number of particles $M = \sum_{x \in \mathbb{T}_L} \eta_x$ is a conserved quantity, but otherwise we assume the process to be irreducible for any fixed M, which is ensured for example by positivity of the rates, i.e. for all $k, l \geq 0$

$$u(k, l) = 0 \quad \Leftrightarrow \quad k = 0 \, .$$

This class includes various models that have been studied in the literature, for example the inclusion process introduced in Giardina et al. (2009), where

$$u(k, l) = k(d + l) \quad \text{for all } k, l \geq 0 \, , \tag{5.27}$$

with a positive parameter $d > 0$. Particles perform independent jumps with rate d and in addition are attracted by each particle on the target site with rate 1, giving rise to the 'inclusion' interaction. This model has attracted recent attention due the presence of condensation phenomena (Grosskinsky et al., 2013; Bianchi et al., 2017) and in the context of large deviations of the particle current (Chleboun et al., 2018). Other well-studied models covered by this set-up are the exclusion process with state space $E \subset \{0, 1\}^{\mathbb{T}_L}$ and $u(\eta_x, \eta_y) = \eta_x(1 - \eta_y)$, or zero-range processes with $E \subset \mathbb{N}_0^{\mathbb{T}_L}$ and rates $u(\eta_x, \eta_y) = u(\eta_x)$ depending only on the occupation number on the departure site.

In terms of previous notation, the jump rates for a lattice gas of type (5.26) between any two configurations η and ζ are given as

$$W(\eta, \zeta) = \sum_{x \in \mathbb{T}_L} \Big(p\, u(\eta_x, \eta_{x+1}) \delta_{\zeta, \sigma_{x,x+1}\eta} + q\, u(\eta_x, \eta_{x-1}) \delta_{\zeta, \sigma_{x,x-1}\eta} \Big) \, . \tag{5.28}$$

In the following we focus on lattice gases where $\sum_x u(\eta_x, \eta_{x+1}) = \sum_x u(\eta_x, \eta_{x-1})$ for all configurations η. While this is not true for models of type (5.26) in general,

it holds for many examples including inclusion, exclusion and zero-range processes mentioned above. With $p+q=1$, the total exit rate out of configuration η is then simply given by

$$\lambda(\eta) = \sum_{x\in\mathbb{T}_L} \Big(p\, u(\eta_x, \eta_{x+1}) + q\, u(\eta_x, \eta_{x-1}) \Big) = \sum_{x\in\mathbb{T}_L} u(\eta_x, \eta_{x+1}) \ . \tag{5.29}$$

In this setting, we are interested in an observable A_T measuring the total particle current up to time T, which is achieved by choosing $h(\eta) \equiv 0$ in (5.1) and

$$g(\eta,\zeta) = \pm 1 \quad \text{if } \zeta = \sigma_{x,x\pm1}\eta \quad \text{and} \quad g(\eta,\zeta) = 0 \quad \text{otherwise} \ .$$

By Lemma 5.2.3, the corresponding tilted process $\mathcal{L}_k$ (5.4) can be written as

$$\mathcal{L}_k(f)(\eta) \ = \ \widehat{\mathcal{L}}_k(f)(\eta) \ + \ \mathcal{V}_k(\eta)\cdot f(\eta) \ ,$$

where the mutation generator is in the form

$$\begin{aligned} \widehat{\mathcal{L}}_k f(\eta) = \sum_{x\in\mathbb{T}_L} \Big[& e^k\, p\, u(\eta_x,\eta_{x+1}) \big(f(\sigma_{x,x+1}\eta) - f(\eta) \big) \\ & + e^{-k}\, q\, u(\eta_x,\eta_{x-1}) \big(f(\sigma_{x,x-1}\eta) - f(\eta) \big) \Big] \ . \end{aligned} \tag{5.30}$$

In particular, the total exit rate out of a configuration η is given by

$$\widehat{\lambda}_k(\eta) \ = \ R_k \sum_{x\in\mathbb{T}_L} u(\eta_x,\, \eta_{x+1}) \ = \ R_k \cdot \lambda(\eta) \ , \quad \text{where } \ R_k := pe^k + qe^{-k} \ ,$$

hence the potential $\mathcal{V}_k$ can be written as

$$\mathcal{V}_k(\eta) \ = \ \widehat{\lambda}_k(\eta) \ - \ \lambda(\eta) \ = \ (R_k - 1)\cdot\lambda(\eta) \ . \tag{5.31}$$

Remark. Note that $R_k = 1$ corresponds to $k = 0$. In this case, $\widehat{\mathcal{L}}_k = \mathcal{L}$, hence the McKean model coincides with the original process and $\mathcal{V}_0 \equiv 0$. In particular, by Proposition 5.2.4, $\Lambda_0 = 0$ and does not require any estimation.

In this setting, we implement two interacting particle approximations to estimate the SCGF associated to the large deviations of the total particle current for inclusion processes. In the following we fix $p, q = 1-p$ and $k \in \mathbb{R}$ and we omit the subscript k for configurations and write $\underline{\eta} = (\eta_i,\ i = 1,\dots,N)$ to simplify notation.

The first algorithm we consider is the cloning algorithm given by (4.31), associated to the McKean model (2.21a) with the common choice $c = 0$, and under

the standard conditions (5.21) and (5.22). Namely, we consider the McKean model characterised by the selection rates

$$\widetilde{W}^{(1)}_{0,k}(\eta,\,\zeta) \;=\; \begin{cases} (R_k-1)\,\lambda(\zeta) & R_k>1 \\ (1-R_k)\,\lambda(\eta) & R_k<1\;. \end{cases}$$

Therefore, if $R_k<1$, the cloning algorithm is given by mutations and killing events which happen independently from each other. However, the similarity between $\widetilde{W}^{(1)}_{0,k}$ and $\widehat{\lambda}_k$ provides a direct relation between mutation and killing rates, and allows us to set up an efficient rejection-based implementation. More precisely, we sample the ensemble of N clones at a total rate of $\lambda^N(\underline{\eta}) := \sum_{i=1}^N \lambda(\eta_i)$, and pick a clone i with probability $\lambda(\eta_i)/\lambda^N(\underline{\eta})$ for the next event. With probability $R_k\in(0,1)$ this is a simple mutation within clone i and we replace η_i with a configuration ζ with probability $\widehat{W}_k(\eta_i,\zeta)/\widehat{\lambda}_k(\eta_i)$, where $\widehat{W}_k$ and $\widehat{\lambda}_k$ are respectively the overall transition rate and the escape rate associated to the mutation generator (5.30). Otherwise, with probability $1-R_k$ we perform a killing event for the clone i and we replace η_i with a configuration η_j, with $j\in\{1,\dots,N\}$ chosen uniformly at random. This procedure ensures that mutation and killing events are sampled with the correct rates.

If $R_k>1$, the cloning algorithm is given by mutations and cloning events, which happen simultaneously with rate $\widehat{\lambda}_k(\eta_i) = R_k\,\lambda(\eta_i)$ for every clone η_i, $i=1,\dots,N$, and, by (5.22), the average number of clones per mutation/cloning event is

$$M_k(\eta_i) \;=\; \frac{R_k-1}{R_k} \in (0,1)\;.$$

In particular, M_k is independent of the state $\underline{\eta}=(\eta_1,\dots,\eta_N)$ of the clone ensemble and, using the common choice (4.25) for the distribution of the number of clones, we see that there is at most one clone per cloning event.

The second algorithm considered is a mean field particle approximation associated to the McKean model given by (2.21b), that is the McKean model characterised by the selection rates

$$\widetilde{W}^{(2)}_{k}(\eta,\,\zeta) \;=\; \begin{cases} (R_k-1)\,\big(\lambda(\zeta)-\lambda(\eta)\big)^+ & R_k>1 \\ (1-R_k)\,\big(\lambda(\eta)-\lambda(\zeta)\big)^+ & R_k<1\;. \end{cases}$$

As for the mutation+killing dynamics described above, the similarity between $\widetilde{W}^{(2)}_{k}$ and $\widehat{\lambda}_k$ provides a direct relation between mutation and selection rates, and allows us to set up an efficient rejection-based implementation for the mean field particle

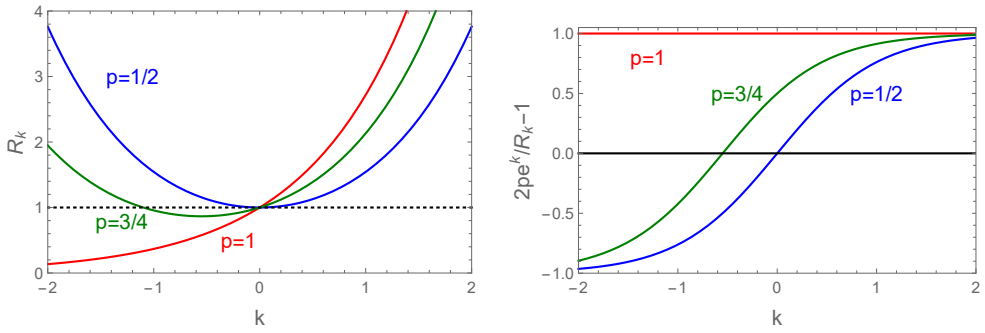

Figure 5.1: Illustration of R_k (left) as given in (5.31) and the drift $2pe^k/R_k - 1$ for the modified dynamics (right) as a function of k for different values of the asymmetry $p = 1 - q$. The minimum of R_k is $2\sqrt{pq}$, attained at $k = \frac{1}{2}\log\frac{q}{p} \in [-\infty, \infty]$, which is also where the modified drift vanishes.

approximation.

In practice, if $R_k < 1$, we sample the ensemble of N clones at a total rate of $\lambda^N(\underline{\eta})$, and pick a clone i with probability $\lambda(\eta_i)/\lambda^N(\underline{\eta})$ for the next event. With probability $R_k \in (0,1)$ this is a simple mutation within clone i, and then we replace η_i with a configuration ζ with probability $\widehat{W}_k(\eta_i,\zeta)/\widehat{\lambda}_k(\eta_i)$. Otherwise, with probability $1 - R_k$ we perform a selection event: we pick a clone $j \in \{1,\dots,N\}$ uniformly at random (including i), if $\lambda(\eta_j) < \lambda(\eta_i)$, we replace η_i by η_j with probability $\big(\lambda(\eta_i) - \lambda(\eta_j)\big)/\lambda(\eta_i)$, otherwise nothing happens.

If $R_k > 1$, we combine the mutation and selection events as in the cloning algorithm and sample the ensemble of N clones at a total rate of $R_k\lambda^N(\underline{\eta})$, and pick a clone i with probability $\lambda(\eta_i)/\lambda^N(\underline{\eta})$ and a clone j uniformly at random. If $\lambda(\eta_j) < \lambda(\eta_i)$, we replace η_j with η_i with probability $\frac{R_k-1}{R_k}\frac{\lambda(\eta_i)-\lambda(\eta_j)}{\lambda(\eta_i)}$, otherwise no selection is performed. Then, we replace η_i with a configuration ζ with probability $\widehat{W}_k(\eta_i,\zeta)/\widehat{\lambda}_k(\eta_i)$.

To avoid confusion, we specify the upper indices (1) and (2) to distinguish the quantities associated respectively to the cloning algorithm with selection rates $\widetilde{W}^{(1)}_{0,k}$ and the mean field particle approximation with selection rates $\widetilde{W}^{(2)}_k$.

Note that, while in both algorithms the total mutation rate is $R_k \cdot \lambda^N(\underline{\eta})$, the selection rates are respectively given by

$$S^{(1)}_k(\underline{\eta}) \;=\; \sum_{i=1}^{N} |\mathcal{V}_k(\eta_i)| \;=\; |R_k - 1|\,\lambda^N(\underline{\eta})\;, \tag{5.32}$$

and

$$S_k^{(2)}(\underline{\eta}) \;=\; \frac{1}{2N}\sum_{i,j=1}^{N}\left|\mathcal{V}_k(\eta_i)-\mathcal{V}_k(\eta_j)\right| \;=\; \frac{|R_k-1|}{2N}\sum_{i,j=1}^{N}\left|\lambda(\eta^i)-\lambda(\eta^j)\right| \le S_k^{(1)}(\underline{\eta})\;. \tag{5.33}$$

So, for very small values of R_k close to 0, the mutation rate can become very small in comparison to selection, which means that significant computation time is devoted to re-weighting by selection, rather than advancing the dynamics via mutation events. This effect is typically much stronger for the standard cloning algorithm with $c = 0$, and occurs for example for totally asymmetric lattice gases with $p = 1$ and negative k conditioning on low currents. In Figure 5.1, we include a sketch of R_k for different values of asymmetry, including also the drift of the modified dynamics, which can be reversed in partially asymmetric systems.

As discussed in Section 4.1.2 and Section 4.2.2, to compare the accuracy of the two interacting particle approximations considered, we need to look at their asymptotic variances, namely $\mathbb{E}\big[V_T^{(1)}(f)^2\big]$ and $\mathbb{E}\big[V_T^{(2)}(f)^2\big]$. Applying the Central Limit Theorem for mean field particle approximations (Proposition 4.1.1) and for cloning algorithms (Proposition 4.2.5), we see that

$$\begin{aligned}\mathbb{E}\big[V_T^{(1)}(f)^2\big] \;-\; \mathbb{E}\big[V_T^{(2)}(f)^2\big] \;=\; \int_0^T \mu_s\Big(&\Gamma_{\widetilde{\mathcal{L}}_{\mu_s}^{(1)}}(\Theta_{s,T}\overline{f},\,\Theta_{s,T}\overline{f}) \;-\; \Gamma_{\widetilde{\mathcal{L}}_{\mu_s}^{(2)}}(\Theta_{s,T}\overline{f},\,\Theta_{s,T}\overline{f})\\ &-\frac{2}{\widehat{\lambda}_k}\,\widehat{\mathcal{L}}_k(\Theta_{s,T}\overline{f})\cdot\widetilde{\mathcal{L}}_{\mu_s}^{(1),t}(\Theta_{s,T}\overline{f})\Big)\,ds\;,\end{aligned}$$

for any $f \in \mathcal{C}_b(E)$, since $Q = 0$, because there is at most one clone per cloning event, as seen above. By Lemma 4.1.4, we know that $\Gamma_{\widetilde{\mathcal{L}}_\mu^{(1)}} - \Gamma_{\widetilde{\mathcal{L}}_\mu^{(2)}}$ is non-negative, for any function $f \in \mathcal{C}_b(E)$ and any probability measure $\mu \in \mathcal{P}(E)$, hence the choice of the McKean model (2.21b) instead of (2.21a) reduces the asymptotic variance. However, we cannot determine a priori the sign of the product of the infinitesimal generators $\widehat{\mathcal{L}}_k(f)\cdot\widetilde{\mathcal{L}}_\mu^{(1),t}(f)$. In other words, the dependence between cloning and mutation events in the cloning algorithm could potentially increase or decrease the total variance, but this also depends on the particular choice of $f \in \mathcal{C}_b(E)$ and $\mu \in \mathcal{P}(E)$.

Considering the very limited analytical results, we want to conclude the section with a numerical comparison of the two algorithms for an inclusion process with $d = 1$, $L = 64$, $M = 128$ and asymmetry $p = 0.7$. It is known (Chleboun et al., 2018) that the SCGF $\Lambda(k)$ scales linearly with the system size L, and outside the convergent regime $k \in [-\ln(\frac{1-p}{p}), 0] \approx [-0.85, 0]$ the rescaled SCGF $\Lambda(k)/L$ diverges as $L \to \infty$ (divergent regime). In Figure 5.2, we compare estimates $\Lambda^{t_0,T,N}(k)$ for

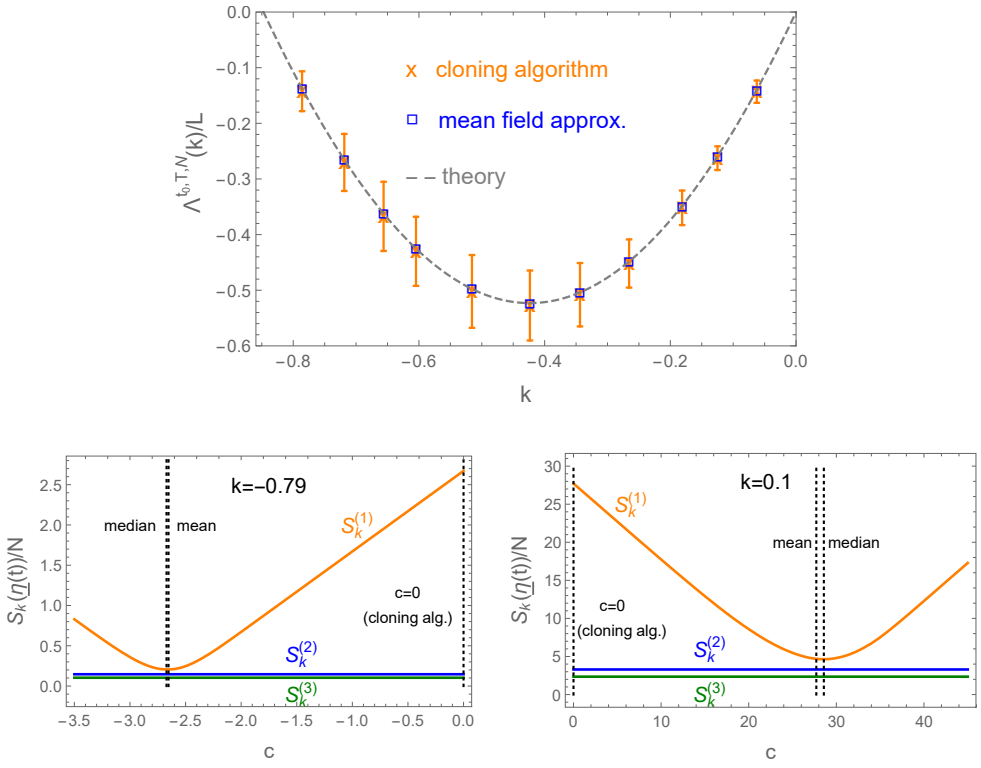

Figure 5.2: Inclusion process (5.27) with $d = 1$, system size $L = 64$, $M = 128$ particles, asymmetry $p = 0.7$ and $N = 2^{11}$ clones at time $t = 42000$. (Top) The rescaled estimator $\Lambda^{t_0,T,N}(k)/L$ as a function of k in the convergent regime, comparing the cloning algorithm with $c = 0$ (orange) and the mean field particle approximation with (2.21b) (blue). Error bars indicate 5 standard deviations, which are bounded by the size of the symbols for the mean field particle approximation. (Bottom) Illustration of the relationship between the total selection rates $S_k^{(1)}$ (depending on c), $S_k^{(2)}$ (5.33) and $S_k^{(3)}$ (5.34) for $k = -0.79$ (left) and $k = 0.1$ (right) based on the state $\underline{\eta}(t)$ of the clone ensemble. Mean and median of the potential $\mathcal{V}_k$ are indicated by dashed vertical lines.

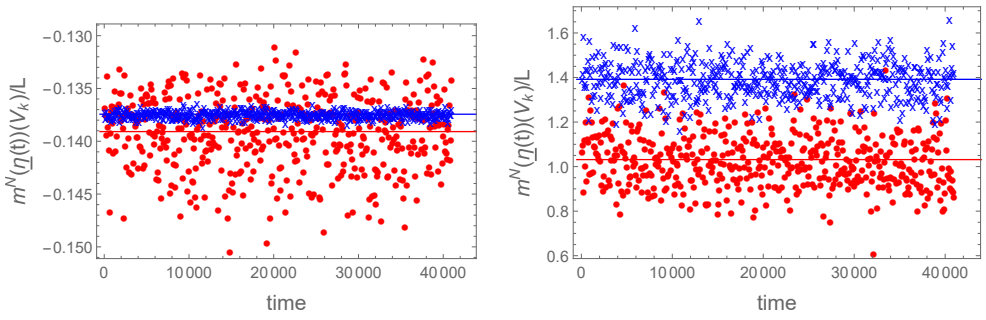

Figure 5.3: Inclusion process (5.27) with $d = 1$, system size $L = 64$, $M = 128$ particles, asymmetry $p = 0.7$ and $N = 2^{11}$ clones. Time series of the mean fitness $m^N(\underline{\eta}(t))(\mathcal{V}_k)/L$ for the cloning algorithm (red dots) and the mean field particle approximation with (2.21b) (blue crosses), with time averages indicated by full lines. (Left) In the convergent regime, namely $[-\ln(\frac{1-p}{p}), 0] \approx [-0.85, 0]$, for $k = -0.79$ we see a clear variance reduction using the mean field particle approximation, but similar time average. (Right) In the divergent regime for $k = 0.1$ we have similar variance but the mean field particle approximation improves on the time average.

the algorithms (1) and (2) in the convergent regime. We use initial conditions where M particles are distributed on L lattice sites uniformly at random, and a burn-in time of $t_0 = 10 \cdot L = 640$, as discussed in Lemma 5.2.5. As we can see from the top of Figure 5.2, both algorithms perform very well and agree with a simple theoretical estimate based on bias reversal, which is not the main concern in this work and we refer the reader to Chleboun et al. (2018). However, enlarged error bars indicating 5 standard deviations reveal that the mean field particle approximation with McKean model (2.21b) is significantly more accurate than the cloning algorithm based on (2.21a). This can be explained by the lower total selection rates, as illustrated at the bottom of Figure 5.2 for the converging and diverging regime. While the total selection rate $S_k^{(2)}$ (5.33) is much lower than $S_k^{(1)}$ with $c = 0$ (5.32), the total selection rate $S_k^{(3)}$ of the McKean model (2.21c), i.e.

$$S_k^{(3)}(\underline{\eta}) \;=\; \frac{1}{2}\sum_{i=1}^{N} \left|\mathcal{V}(x_i) - m(\underline{x})(\mathcal{V})\right| \;\leq\; S_k^{(2)}(\underline{\eta}) \;, \tag{5.34}$$

does not offer significant further improvement. Since the efficient rejection based implementation of the mean field particle approximation explained above does not work for the McKean model (2.21c), we focus on the McKean model (2.21b) in our context.

As illustrated in Figure 5.3, the cloning algorithm with (2.21a) has a signifi-

cantly higher time variation of the average potential $m^N(\underline{\eta}(t))(\mathcal{V}_k)/L$ in the convergent regime compared to the mean field particle approximation with (2.21b). Hence, in comparison to the classical cloning, the mean field particle approximation leads to reduced finite size effects and/or a significant variance reduction in this example, and a significant improvement of convergence of the estimator $\Lambda^{t_0,T,N}(k)$. This is due to the much higher selection rate for the McKean model (2.21a) and possibly to the dependence between mutation and cloning events in the cloning algorithm. These promising first numerical results pose interesting questions for a systematic study of practical properties of the algorithms and associated time correlations for future work, also in comparison with various recent results on improvements of cloning algorithms (Nemoto et al., 2016; Brewer et al., 2018; Ferré and Touchette, 2018).

Chapter 6

Conclusion and Outlook

We conclude with a summary of the main contributions of this work, mentioning a couple of interesting directions for future works.

In Chapter 3, we adapted already established convergence results for mean field approximations (Del Moral and Miclo, 2000; Rousset, 2006) in the context of pure jump processes to a broader class of interacting particle approximations, providing general assumptions on the infinitesimal generator (condition 3.3a), carré du champ (condition 3.3b) and jumps (condition 3.3c) of the interacting particle system. These guarantee that the empirical distribution μ_t^N converges uniformly in time to μ_t in L^p and almost surely in the weak topology, with order of convergence of the L^p and systematic errors given respectively by $1/\sqrt{N}$ and $1/N$ (see Theorem 3.1.2), as for mean field approximations. We also provided a central limit theorem (Theorem 3.1.3) and explicit formulas for the asymptotic variance of the algorithms. These results apply in the general setting of jump Markov processes on locally compact state spaces. Essential conditions for this approach are summarised in Assumption 2.2.2 on asymptotic stability of the Feynman-Kac model and Assumption 3.1.1 on the particle approximation, which are usually straightforward to check for practical applications. The proofs are based on the martingale characterisation of the particle system and also make use of the propagator $\Theta_{t,T}$ and its differential equation. These results underline the several degrees of freedom in the design of the algorithms, providing a new perspective on how to optimise the implementation of interacting particle approximations. Our results apply in the general setting of jump Markov processes on locally compact state spaces. A future development would be the adaptation of similar statistical techniques to more general processes, including piecewise deterministic processes (Davis, 1984; Finke et al., 2014), jump-diffusions,

or possibly non-Markovian dynamics (see e.g. Cavallaro and Harris, 2016, for first heuristic results in this direction).

In Chapter 4, we described the construction of mean field particle approximations, cloning algorithms and a particular version of resampling algorithms, and discussed the limit properties of these methods. The main contribution of the Chapter is the novel interpretation of the cloning algorithm based on Feynman-Kac models and the infinitesimal description of the algorithm through its generator and associated carré du champ. Once we obtain an explicit formulation for these two quantities, it is then easy to check that Assumption 3.1.1 is satisfied for the cloning algorithm and, hence, that Theorem 3.1.2 and Theorem 3.1.3 hold. This allows us to establish rigorous error bounds for the cloning algorithm in continuous time, and to suggest different versions of the algorithm based on different McKean interpretations of Feynman-Kac models.

Finally, in Chapter 5, we show how to apply interacting particle approximations in the study of large deviations for additive path functions of stochastic processes. In particular, by interpreting the scaled cumulant generating function (SCGF) via Feynman-Kac measures, we can construct an estimator $\Lambda^{t,T,N}(k)$ for the SCGF based on the empirical distribution of an interacting particle system and show that $\Lambda^{t,T,N}(k)$ converges in L^p to the finite-time approximation $\Lambda^{t,T}(k)$ of the SCGF, with order of convergence of the bias and L^p error given by $1/\sqrt{N}$. In Section 5.4 we discuss another estimator for $\Lambda^{t,T}(k)$ based on the cloning factor, which is commonly used in the physics literature (we recall, for instance, Giardinà et al., 2006; Pérez-Espigares and Hurtado, 2019). Our contribution is the analytical justification of this estimator by showing that it converges to $\Lambda^{t,T}(k)$ as the size N of the system goes to infinity and the rate of convergence is still $1/\sqrt{N}$, as for the first estimator. The proof is based on the martingale characterisation of the cloning factor and on its connection with the first estimator $\Lambda^{t,T,N}(k)$.

An outlook on asymptotic variance estimators. As discussed in Section 4.1.2 and Section 4.2.2, in order to study the accuracy of an interacting particle approximation, we need to look at the asymptotic variance $\mathbb{E}\big[V_T(f)^2\big]$, which unfortunately cannot be estimated a priori, as it depends on μ_t and on the corresponding propagator $\Theta_{t,T}$, $t \in [0,T]$. An interesting direction for future research would be the construction of an estimator of the asymptotic variance based on a single realisation of an interacting particle approximation. Recent results for discrete-time Markov

processes by Lee and Whiteley (2018) and Du and Guyader (2019) are based on coalescent tree-based measures. However, most of the tools used in the discrete-time case fail to work in the continuous-time. For instance, the explicit expression of the asymptotic variance (3.8), that in our case is an integral over $[0, T]$, namely

$$\mathbb{E}\big[V_T(f)^2\big] = \mu_0\big((\Theta_{0,T}\overline{f})^2\big) \; + \; \int_0^T \mu_s\Big(G_{\mu_s}\big(\Theta_{s,T}\overline{f},\; \Theta_{s,T}\overline{f}\big)\Big)\, ds \;,$$

in the discrete-time case is a simple summation over the times $p = 1, \ldots, n$. One of the key ideas in the work of Du and Guyader (2019) is to interpret every term in the summation as a coalescent tree-based measure with only one coalescence at time p. This of course cannot work in the continuous-time setting, however it could be possible to define in an appropriate way a random sequence of times τ_p and then consider a coalescent tree-based measure with only a coalescence at time τ_p, adapt the results by Du and Guyader (2019) and then conclude by Monte Carlo integration. The work of Lee and Whiteley (2018) provides a different insight based on tensor product measures that could be more natural to adapt in the continuous-time case.

An outlook on driven process estimators. In the context of interacting particle approximations for the study of large deviations, one interesting direction for further research would be the study of the connection between McKean models and the *driven process* (also known in the literature as Doob's h-transform or twisted Markov kernel), which is a key object in the study of large deviations in statistical mechanics (see, for example, Garrahan, 2016; Oakes et al., 2018; Chetrite and Touchette, 2015). We mentioned it in Section 2.4, namely in Equation (2.22), and we recall here that the h-transform of a tilted generator $\mathcal{L}_k$ is the infinitesimal generator L_k defined as

$$L_k := r_k^{-1}\mathcal{L}_k r_k - r_k^{-1}(\mathcal{L}_k r_k) \;=\; r_k^{-1}\mathcal{L}_k r_k - \Lambda(k)\,,$$

where r_k is the right eigenfunction associated to the principal eigenvalue $\Lambda(k)$ of $\mathcal{L}_k$. Although the tilted generator $\mathcal{L}_k$ is not conservative, L_k is, since $(L_k 1) = 0$, and it remains a pure jump process described by the modified kernel

$$W_k(x, dy) = r_k^{-1}(x) W(x, dy) e^{kg(x,y)} r_k(y),$$

for all $x, y \in E$. Unfortunately, the right eigenfunction cannot be determined analytically, so numerical procedures are required to estimate r_k, and thus the kernel of the driven process (see e.g. Ferré and Touchette, 2018). Interpreting r_k via Feynman-Kac

measures, as we have done for the principal eigenvalue $\Lambda(k)$ and the left eigenmeasure μ_∞, could provide an estimator for r_k based on interacting particle approximations. The idea is to adapt some previous results for discrete-time Feynman-Kac models (Whiteley and Lee, 2014; Whiteley and Kantas, 2017) to the continuous-time case. We briefly illustrate here some preliminary discussion about this approach in the continuous-time case.

First, recall that, in Section 5.2, we started by providing a 'deterministic' approximation of the eigenquantities $\Lambda(k)$ and μ_∞ given by $\Lambda^{aT,T}(k)$ (5.16) and μ_T, respectively, and then we estimated these two quantities numerically via interacting particle approximations. In a similar fashion, we want to first approximate r_k via a sequence of functions $\big(r_{s,t}\big)_{0\le s<t\le\infty}$, with $r_{s,t} \in \mathcal{C}_b(E)$, and then approximating these via numerical procedures. Thus, a candidate for estimating the h-transform L_k is given by

$$L_{s,t}(f)(x) \;=\; r_{s,t}^{-1}(x)\cdot\mathcal{L}_k(r_{s,t}f)(x) \;-\; r_{s,t}^{-1}(x)\cdot\mathcal{L}_k(r_{s,t})(x)\cdot f(x)\,,$$

for every $f \in \mathcal{C}_b(E)$.

Mimicking the discrete-time case (Whiteley and Lee, 2014, Section 2.3), a good candidate is given by

$$r_{s,t}(x) := \frac{\nu_{t-s,x}(1)}{\nu_{t-s,\mu_s}(1)}\,,$$

for every $x \in E$ and $s \le t$, where $\nu_{t,x}$ and $\nu_{t,\mu}$ are the unnormalized t-marginal Feynman-Kac measures (associated to the tilted generator $\mathcal{L}_k$) with initial distribution respectively given by δ_x and μ. Note that, by definition, $r_{t,t}(x) = 1$ and $\mu_s(r_{s,t}) = 1$. Moreover, the functions $r_{s,t}$ satisfy a recursive relationship that resembles the one in the discrete-time case (see Whiteley and Lee, 2014, Lemma 1), as illustrated by the following result.

Lemma 6.1. *The function $s \mapsto r_{s,t}$ is left-differentiable and the left-derivative is given by*

$$\frac{d_-}{ds}r_{s,t}(x) \;=\; -\,\mathcal{L}_k(r_{s,t})(x) \;+\; \mu_s(\mathcal{V}_k)\cdot r_{s,t}(x)\,.$$

Remark. Compare the backward recursion for $r_{s,t}$,

$$\frac{d_-}{ds}r_{s,t}(x) \;=\; -\,\widehat{\mathcal{L}}_k(r_{s,t})(x) \;-\; \mathcal{V}_k(x)\,r_{s,t}(x) \;+\; \mu_s(\mathcal{V}_k)\cdot r_{s,t}(x)\;,$$

with the forward equation for μ_t (2.2.4), namely

$$\frac{d}{dt}\mu_t(f) \;=\; \mu_t\big(\widehat{\mathcal{L}}_k(f)\big) \;-\; \mu_t(\mathcal{V}_k\,f) \;+\; \mu_t(\mathcal{V}_k)\cdot\mu_t(f)\;,\quad f\in\mathcal{C}_b(E)\;.$$

Proof. By the Markov property of $\nu_{t,x}$, we have

$$\begin{aligned}\nu_{t-s,x}(1) &= \int_E \nu_{\varepsilon,x}(dz)\cdot \nu_{t-s-\varepsilon,z}(1)\\ &= \int_E \nu_{\varepsilon,x}(dz)\cdot r_{s+\varepsilon,t}(z)\cdot \nu_{t-s-\varepsilon,\mu_s}(1)\\ &= \nu_{t-s-\varepsilon,\mu_s}(1)\cdot \nu_{\varepsilon,x}\big(r_{s+\varepsilon,t}\big)\ ,\end{aligned}$$

for every $0\le\varepsilon\le t-s$. Moreover,

$$\frac{d}{dt}\log\nu_{t,\mu_0}(1) = \frac{1}{\nu_{t,\mu_0}(1)}\cdot\frac{d}{dt}\nu_{t,\mu_0}(1) = \frac{\nu_{t,\mu_0}(\mathcal{L}_k(1))}{\nu_{t,\mu_0}(1)} = \mu_t(\mathcal{V}_k).$$

And, thus,

$$\nu_{t,\mu_0}(1) = \exp\bigg(\int_0^t \mu_s(\mathcal{V}_k)\,ds\bigg),$$

since $\nu_{0,\mu_0}(1)=1$.

Thanks to these considerations, we can rewrite $r_{s,t}$ as

$$r_{s,t}(x) = \exp\bigg(-\int_s^{s+\varepsilon}\mu_u(\mathcal{V}_k)\,du\bigg)\cdot\nu_{\varepsilon,x}\big(r_{s+\varepsilon,t}\big)\,,$$

for every $0\le\varepsilon\le t-s$. Therefore,

$$\begin{aligned}\frac{d_-}{ds}\,r_{s,t}(x) &= \lim_{\varepsilon\to0_+}\frac{r_{s,t}(x)\,-\,r_{s-\varepsilon,t}(x)}{\varepsilon}\\ &= \lim_{\varepsilon\to0_+}\frac{r_{s,t}(x)\,-\,\exp\Big(-\int_{s-\varepsilon}^s\mu_u(\mathcal{V}_k)\,du\Big)\cdot\nu_{\varepsilon,x}\big(r_{s,x}\big)}{\varepsilon}\\ &= \lim_{\varepsilon\to0_+} r_{s,t}(x)\cdot\frac{1-\exp\Big(-\int_{s-\varepsilon}^s\mu_u(\mathcal{V}_k)\,du\Big)}{\varepsilon}\\ &\qquad-\lim_{\varepsilon\to0_+}\exp\bigg(-\int_{s-\varepsilon}^s\mu_u(\mathcal{V}_k)\,du\bigg)\cdot\frac{\nu_{\varepsilon,x}(r_{s,t})-r_{s,t}(x)}{\varepsilon}\ .\end{aligned}$$

The conclusion follows by observing that

$$\begin{aligned}&\lim_{\varepsilon\to0_+} r_{s,t}(x)\cdot\frac{1-\exp\Big(-\int_{s-\varepsilon}^s\mu_u(\mathcal{V}_k)\,du\Big)}{\varepsilon}\\ &\qquad= r_{s,t}(x)\,\exp\Big(-\int_0^s\mu_u(\mathcal{V}_k)\,du\Big)\,\frac{d_-}{ds}\exp\Big(\int_0^s\mu_u(\mathcal{V}_k)\,du\Big)\\ &\qquad= r_{s,t}(x)\,\mu_s(\mathcal{V}_k)\ ,\end{aligned}$$

and

$$\lim_{\varepsilon \to 0_+} \exp\left(- \int_{s-\varepsilon}^{s} \mu_u(\mathcal{V}_k)\, du \right) \cdot \frac{\nu_{\varepsilon,x}(r_{s,t}) - r_{s,t}(x)}{\varepsilon} = \frac{d}{du} \nu_{u,x}(r_{s,t})\big|_{u=0} = \mathcal{L}_k(r_{s,t})(x) \ .$$

□

This preliminary result suggests that the results on exponential convergence for the discrete-time case (see Whiteley and Lee, 2014, Proposition 1) could be adapted to the continuous-time case and we expect that, for any $0 \leq s \leq t$, there exist constants α, $\beta > 0$ and $\rho \in (0,1)$ (independent of the initial distribution μ_0) such that

$$\|r_{s,t} - r_k\| \leq \alpha\, \rho^{(t-s)\wedge s} \ ,$$
$$\|L_{s,t}(f) - L_k(f)\| \leq \|f\| \cdot \beta\, \rho^{(t-s)\wedge s} \ ,$$

for every $f \in \mathcal{C}_b(E)$. The backward equation given by Lemma 6.1 also suggests that $r_{s,t}$ could be estimated via a forward-backward particle approximation, first running the interacting particle system forward till time $2T$ and then proceeding backward in an appropriate way, from time $2T$ to time T. However, this construction would be of little practicality in terms of applications, and one way to improve it is to update the driven kernel in time. One approach that could be worthwhile inspecting is the following. Considering a (possibly random) sequence of times $0 = t_0 < t_1 < t_2 < \ldots$, such that $t_n - t_{n-1}$ diverges as $n \to \infty$, we construct recursively a sequence of functions $(h_n)_{n\in\mathbb{N}}$, $h_n \in \mathcal{C}_b(E)$ and a sequence of generators $(L_n)_{n\in\mathbb{N}}$ respectively by

$$h_1 = r_{0,t_1} \ , \quad h_n = r_{t_{n-1},t_n} \cdot h_{n-1}^{-1} \ ,$$

and

$$L_0 = \mathcal{L}_k \ , \quad L_n = h_n^{-1} L_{n-1}(h_n) \ ,$$

for any $n \in \mathbb{N}$, $n > 1$. Therefore, we see that $r_{t_{n-1},t_n} = h_n \ldots h_1$ and $L_{t_{n-1},t_n} = L_n$. This would give a forward in time approximation of the driven process generator.

Bibliography

J. B. Anderson. A random-walk simulation of the Schrödinger equation: H_3^+. *The Journal of Chemical Physics*, 63(4):1499–1503, 1975.

L. Angeli, S. Grosskinsky, A. M. Johansen, and A. Pizzoferrato. Rare event simulation for stochastic dynamics in continuous time. *Journal of Statistical Physics*, 176(5):1185–1210, 2019.

L. Angeli, S. Grosskinsky, and A. M. Johansen. Limit theorems for cloning algorithms. *arXiv preprint arXiv:1902.00509*, 2020.

J. E. Baker. Adaptive selection methods for genetic algorithms. In *Proceedings of an International Conference on Genetic Algorithms and their applications*, pages 101–111. Hillsdale, New Jersey, 1985.

L. Bertini, A. Faggionato, and D. Gabrielli. Large deviations of the empirical flow for continuous time Markov chains. In *Annales de l'IHP Probabilités et statistiques*, volume 51, pages 867–900, 2015.

A. Bianchi, S. Dommers, and C. Giardinà. Metastability in the reversible inclusion process. *Electronic Journal of Probability*, 22, 2017.

Y. V. Borovskikh and V. Korolyuk. *Martingale approximation*. Walter de Gruyter GmbH & Co KG, 2019.

T. Brewer, S. R. Clark, R. Bradford, and R. L. Jack. Efficient characterisation of large deviations using population dynamics. *Journal of Statistical Mechanics: Theory and Experiment*, 2018(5):053204, 2018.

M. Cavallaro and R. J. Harris. A framework for the direct evaluation of large deviations in non-Markovian processes. *Journal of Physics A: Mathematical and Theoretical*, 49(47):47LT02, 2016.

F. Cérou, B. Delyon, A. Guyader, and M. Rousset. A central limit theorem for Fleming-Viot particle systems with soft killing. *arXiv preprint arXiv:1611.00515*, 2016.

R. Chetrite and H. Touchette. Nonequilibrium Markov processes conditioned on large deviations. *Annales Henri Poincaré*, 16:2005–2057, 2015.

P. Chleboun, S. Grosskinsky, and A. Pizzoferrato. Current large deviations for partially asymmetric particle systems on a ring. *Journal of Physics A: Mathematical and Theoretical*, 51(40):405001, 2018.

H. Cramér. *Mathematical methods of statistics*, volume 43. Princeton university press, 1999.

M. H. Davis. Piecewise-deterministic Markov processes: a general class of non-diffusion stochastic models. *Journal of the Royal Statistical Society: Series B (Methodological)*, 46(3):353–376, 1984.

P. Del Moral. *Feynman-Kac formulae: genealogical and interacting particle systems with applications*. Springer Science & Business Media, 2004.

P. Del Moral. *Mean field simulation for Monte Carlo integration*. CRC Press, 2013.

P. Del Moral and L. Miclo. Branching and interacting particle systems approximations of Feynman-Kac formulae with applications to non-linear filtering. In *Seminaire de probabilites XXXIV*, pages 1–145. Springer, 2000.

P. Del Moral and L. Miclo. Particle approximations of Lyapunov exponents connected to Schrödinger operators and Feynman-Kac semigroups. *ESAIM: Probability and Statistics*, 7:171–208, 2003.

P. Del Moral and L. Miclo. Strong propagations of chaos in Moran's type particle interpretations of Feynman–Kac measures. *Stochastic analysis and applications*, 25(3):519–575, 2007.

A. Dembo and O. Zeitouni. *Large deviations techniques and applications*, volume 38. Springer Science & Business Media, 2009.

F. Den Hollander. *Large deviations*, volume 14. American Mathematical Soc., 2008.

R. Douc, O. Cappé, and E. Moulines. Comparison of resampling schemes for particle filtering. In *ISPA 2005. Proceedings of the 4th International Symposium on Image and Signal Processing and Analysis*, pages 64–69. IEEE, 2005.

A. Doucet and A. M. Johansen. A tutorial on particle filtering and smoothing: Fifteen years later. *The Oxford Handbook of Nonlinear Filtering*, D. Crisan and B. Rozovsky (eds.):656–704, 2011.

D. Down, S. P. Meyn, and R. L. Tweedie. Exponential and uniform ergodicity of Markov processes. *Annals of Probability*, pages 1671–1691, 1995.

Q. Du and A. Guyader. Variance estimation in adaptive sequential Monte Carlo. *arXiv preprint arXiv:1909.13602*, 2019.

S. N. Ethier and T. G. Kurtz. *Markov processes: characterization and convergence*, volume 282. John Wiley & Sons, 2009.

E. A. Feinberg, P. O. Kasyanov, and Y. Liang. Fatou's lemma for weakly converging measures under the uniform integrability condition. *Theory of Probability & Its Applications*, 64(4):615–630, 2020.

G. Ferré and H. Touchette. Adaptive sampling of large deviations. *Journal of Statistical Physics*, 172(6):1525–1544, 2018.

R. P. Feynman. Space-time approach to quantum electrodynamics. *Physical Review*, 76(6):769, 1949.

A. Finke, A. M. Johansen, and D. Spanò. Static-parameter estimation in piecewise deterministic processes using particle Gibbs samplers. *Annals of the Institute of Statistical Mathematics*, 66(3):577–609, 2014.

J. P. Garrahan. Classical stochastic dynamics and continuous matrix product states: gauge transformations, conditioned and driven processes, and equivalence of trajectory ensembles. *Journal of Statistical Mechanics: Theory and Experiment*, 2016 (7):073208, 2016.

A. Gelman and K. Shirley. Inference from simulations and monitoring convergence. *Handbook of markov chain monte carlo*, 6:163–174, 2011.

M. Gerber, N. Chopin, and N. Whiteley. Negative association, ordering and convergence of resampling methods. *Annals of Statistics*, 47(4):2236–2260, 2019.

C. Giardinà, J. Kurchan, and L. Peliti. Direct evaluation of large-deviation functions. *Physical Review Letters*, 96(12):120603, 2006.

C. Giardina, J. Kurchan, F. Redig, and K. Vafayi. Duality and hidden symmetries in interacting particle systems. *Journal of Statistical Physics*, 135(1):25–55, 2009.

C. Giardinà, J. Kurchan, V. Lecomte, and J. Tailleur. Simulating rare events in dynamical processes. *Journal of Statistical Physics*, 145(4):787–811, 2011.

C. Giardinà, C. Giberti, and E. Magnanini. Approximating the cumulant generating function of triangles in the Erdös-Rényi random graph. *arXiv preprint arXiv:2007.12971*, 2020.

F. Gong and L. Wu. Spectral gap of positive operators and applications. *Journal de Mathématiques Pures et Appliquées*, 85(2):151–191, 2006.

F. Gong, M. Röckner, and L. Wu. Poincaré inequality for weighted first order Sobolev spaces on loop spaces. *Journal of Functional Analysis*, 185(2):527–563, 2001.

P. Grassberger. Go with the winners: A general Monte Carlo strategy. *Computer Physics Communications*, 147(1-2):64–70, 2002.

S. Grosskinsky, F. Redig, and K. Vafayi. Dynamics of condensation in the symmetric inclusion process. *Electronic Journal of Probability*, 18, 2013.

K. Hamza and F. C. Klebaner. Conditions for integrability of Markov chains. *Journal of Applied Probability*, 32(2):541–547, 1995.

E. G. Hidalgo, T. Nemoto, and V. Lecomte. Finite-time and finite-size scalings in the evaluation of large-deviation functions: Numerical approach in continuous time. *Physical Review E*, 95(6):062134, 2017.

J. D. Hol, T. B. Schon, and F. Gustafsson. On resampling algorithms for particle filters. In *2006 IEEE nonlinear statistical signal processing workshop*, pages 79–82. IEEE, 2006.

P. I. Hurtado and P. L. Garrido. Test of the additivity principle for current fluctuations in a model of heat conduction. *Physical Review Letters*, 102(25):250601, 2009.

P. I. Hurtado, C. P. Espigares, J. J. del Pozo, and P. L. Garrido. Thermodynamics of currents in nonequilibrium diffusive systems: theory and simulation. *Journal of Statistical Physics*, 154(1-2):214–264, 2014.

J. Jacod and A. Shiryaev. *Limit theorems for stochastic processes*, volume 288. Springer Science & Business Media, 2013.

M. Kac. On distributions of certain Wiener functionals. *Transactions of the American Mathematical Society*, 65(1):1–13, 1949.

A. Kong, J. S. Liu, and W. H. Wong. Sequential imputations and Bayesian missing data problems. *Journal of the American Statistical Association*, 89(425):278–288, 1994.

R. Kutner and J. Masoliver. The continuous time random walk, still trendy: fifty-year history, state of art and outlook. *The European Physical Journal B*, 90(3): 50, 2017.

J. L. Lebowitz and H. Spohn. A Gallavotti-Cohen-type symmetry in the large deviation functional for stochastic dynamics. *Journal of Statistical Physics*, 95(1): 333–365, 1999.

V. Lecomte and J. Tailleur. A numerical approach to large deviations in continuous time. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(03):P03004, 2007.

A. Lee and N. Whiteley. Variance estimation in the particle filter. *Biometrika*, 105 (3):609–625, 2018.

T. Lestang, F. Bouchet, and E. Lévêque. Numerical study of extreme mechanical force exerted by a turbulent flow on a bluff body by direct and rare-event sampling techniques. *Journal of Fluid Mechanics*, 895, 2020.

T. M. Liggett. *Continuous time Markov processes: an introduction*, volume 113. American Mathematical Soc., 2010.

T. M. Liggett. *Interacting particle systems*, volume 276. Springer Science & Business Media, 2012.

T. Nemoto, F. Bouchet, R. L. Jack, and V. Lecomte. Population-dynamics method with a multicanonical feedback control. *Physical Review E*, 93(6):062123, 2016.

T. Nemoto, E. G. Hidalgo, and V. Lecomte. Finite-time and finite-size scalings in the evaluation of large-deviation functions: Analytical study using a birth-death process. *Physical Review E*, 95(1):012102, 2017.

T. Nemoto, É. Fodor, M. E. Cates, R. L. Jack, and J. Tailleur. Optimizing active work: Dynamical phase transitions, collective motion, and jamming. *Physical Review E*, 99(2):022605, 2019.

T. Oakes, S. Powell, C. Castelnovo, A. Lamacraft, and J. P. Garrahan. Phases of quantum dimers from ensembles of classical stochastic trajectories. *Physical Review B*, 98(6):064302, 2018.

C. Pérez-Espigares and P. I. Hurtado. Sampling rare events across dynamical phase transitions. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(8): 083106, 2019.

E. Pitard, V. Lecomte, and F. Van Wijland. Dynamic transition in an atomic glass former: A molecular-dynamics evidence. *EPL (Europhysics Letters)*, 96(5):56002, 2011.

F. Ragone, J. Wouters, and F. Bouchet. Computation of extreme heat waves in climate models using a large deviation algorithm. *Proceedings of the National Academy of Sciences*, 115(1):24–29, 2018.

M. Rousset. On the control of an interacting particle estimation of Schrödinger ground states. *SIAM Journal on Mathematical Analysis*, 38(3):824–844, 2006.

S. M. Schmon, G. Deligiannidis, A. Doucet, and M. K. Pitt. Large sample asymptotics of the pseudo-marginal method. *Biometrika*, 2020. In press.

R. Serfozo. Convergence of Lebesgue integrals with varying measures. *Sankhyā: The Indian Journal of Statistics, Series A*, 44(3):380–402, 1982.

M. Tchernookov and A. R. Dinner. A list-based algorithm for evaluation of large deviation functions. *Journal of Statistical Mechanics: Theory and Experiment*, 2010(02):P02006, 2010.

H. Touchette. *Equivalence and nonequivalence of the microcanonical and canonical ensembles: a large deviations study.* PhD thesis, McGill University, 2003.

H. Touchette. The large deviation approach to statistical mechanics. *Physics Reports*, 478(1):1–69, 2009.

R. L. Tweedie. Topological conditions enabling use of Harris methods in discrete and continuous time. *Acta Applicandae Mathematica*, 34(1-2):175–188, 1994.

N. Whiteley and N. Kantas. Calculating principal eigen-functions of non-negative integral kernels: particle approximations and applications. *Mathematics of Operations Research*, 42(4):1007–1034, 2017.

N. Whiteley and A. Lee. Twisted particle filters. *The Annals of Statistics*, 42(1): 115–141, 2014.